SEPTEMBER/OCTOBER 2015 • VOLUME 86, NUMBER 5

CHILD DEVELOPMENT

Editor

CYNTHIA GARCIA COLL
Carlos Albizu University

Associate Editors

Amy E. Booth, *University of Texas at Austin*
J. Gavin Bremner, *Lancaster University*
Natasha J. Cabrera, *University of Maryland*
Carol McDonald Connor, *Arizona State University*
Michael Cunningham, *Tulane University*
David K. Dickinson, *Vanderbilt University*
Anne Fernald, *Stanford University*
Lisa A. Gennetian, *New York University*
Gustaf Gredebäck, *Uppsala University*
Amanda E. Guyer, *University of California, Davis*
Marinus van IJzendoorn, *Leiden University, Erasmus University Rotterdam*
Bonnie J. Leadbeater, *University of Victoria, B.C. Canada*
Nonie K. Lesaux, *Harvard University*
Katherine A. Magnuson, *University of Wisconsin, Madison*
Tina Malti, *University of Toronto*
Diane Poulin-Dubois, *Concordia University*
Nirmala Rao, *The University of Hong Kong*
Glenn I. Roisman, *University of Minnesota*
Megan Saylor, *Vanderbilt University*
Carlo Schuengel, *VU University Amsterdam*
Virginia Slaughter, *University of Queensland*
Charles M. Super, *University of Connecticut*
Jennifer Brown Urban, *Montclair State University*

CELEBRATING
1930-2015
85th
Anniversary
CHILD DEVELOPMENT

SRCD
Society for Research in Child Development

Marygrove College Library
8425 West McNichols Road
Detroit, MI 48221

Published by Wiley for the Society for Research in Child Development

Published by Wiley for the Society for Research in Child Development

Child Development

Disclaimer

The Publisher, The Society for Research in Child Development, Inc. and Editors cannot be held responsible for errors or any consequences arising from the use of information contained in this journal; the views and opinions expressed do not necessarily reflect those of the Publisher, The Society for Research in Child Development, Inc. and Editors, neither does the publication of advertisements constitute any endorsement by the Publisher, The Society for Research in Child Development, Inc. and Editors of the products advertised.

Copyright and Copying

Child Development © 2015 The Society for Research in Child Development, Inc. All rights reserved. No part of this publication may be reproduced, stored or transmitted in any form or by any means without the prior permission in writing from the copyright holder. Authorization to copy items for internal and personal use is granted by the copyright holder for libraries and other users registered with their local Reproduction Rights Organisation (RRO), e.g. Copyright Clearance Center (CCC), 222 Rosewood Drive, Danvers, MA 01923, USA (www.copyright.com), provided the appropriate fee is paid directly to the RRO. This consent does not extend to other kinds of copying such as copying for general distribution, for advertising or promotional purposes, for creating new collective works or for resale. Special requests should be addressed to: permissionsuk@wiley.com

Information for subscribers

Child Development is published in 6 issues per year. Institutional subscription prices for Child Development/Child Development Perspectives/Monographs of the Society for Research in Child for 2015 are: Print & Online: US$987 (USA), US$1380 (Rest of World), €897 (Europe), £706 (UK). Prices are exclusive of tax. Asia-Pacific GST, Canadian GST/HST and European VAT will be applied at the appropriate rates. For more information on current tax rates, please go to www.wileyonlinelibrary.com/tax-vat. The price includes online access to the current and all online back files to January 1st 2011, where available. For other pricing options, including access information and terms and conditions, please visit www.wileyonlinelibrary.com/access.

Delivery Terms and Legal Title

Where the subscription price includes print issues and delivery is to the recipient's address, delivery terms are Delivered at Place (DAP); the recipient is responsible for paying any import duty or taxes. Title to all issues transfers FOB our shipping point, freight prepaid. We will endeavour to fulfil claims for missing or damaged copies within six months of publication, within our reasonable discretion and subject to availability.

Back issues: Single issues from current and recent volumes are available at the current single issue price from cs-journals@wiley.com. Earlier issues may be obtained from Periodicals Service Company, 351 Fairview Avenue – Ste 300, Hudson, NY 12534, USA. Tel: +1 518 822-9300, Fax: +1 518 822-9305, Email: psc@periodicals.com

CHILD DEVELOPMENT (ISSN 0009-3920), is published bimonthly, on behalf of Society for Research in Child Development, Inc., by Wiley Subscription Services, Inc., a Wiley Company, 111 River St., Hoboken, NJ 07030-5774.

Application to mail at the periodical rate is pending at Hoboken, NJ and additional offices.

Postmaster: Send all address changes to CHILD DEVELOPMENT, John Wiley & Sons Inc., C/O The Sheridan Press, PO Box 465, Hanover, PA 17331.

Journal Customer Services: For ordering information, claims and any enquiry concerning your journal subscription please go to www.wileycustomerhelp.com/ask or contact your nearest office.

Americas: Email: cs-journals@wiley.com; Tel: +1 781 388 8598 or +1 800 835 6770 (toll free in the USA & Canada).

Europe, Middle East and Africa: Email: cs-journals@wiley.com; Tel: +44 (0) 1865 778315.

Asia Pacific: Email: cs-journals@wiley.com; Tel: +65 6511 8000.

Japan: For Japanese speaking support, Email: cs-japan@wiley.com; Tel: +65 6511 8010 or Tel (toll-free): 005 316 50 480.

Visit our Online Customer Get-Help available in 6 languages at www.wileycustomerhelp.com

Production Editor: Kate Noe (CDEV@wiley.com)

Advertising: Kristin McCarthy (kmccarthy@wiley.com)

Indexing/Abstracting: The contents of this journal are indexed or abstracted in the following: *Academic ASAP; Expanded Academic ASAP; Academic Search Elite; Academic Search Premier; Applied Social Sciences Index and Abstracts; Biomedical Reference Collection: Comprehensive Edition; CAB Abstracts; CAB Direct; Contents Pages in Education; Current Contents/Social & Behavioral Science; Current Index to Journals in Education; Cambridge Scientific Abstracts Linguistics & Language Behavior Abstracts; CatchWord; Current Index to Journals in Education; Educational Research Abstracts; Environmental Sciences & Pollution Management; e-psyche; EBSCO Sociological Collection; EBSCO Online; EBSCO Professional Development Collection; Educational Research Abstracts; Educational Resources Information Center; Health and Safety Science Abstracts; ISI Basic Social Sciences Index; InfoTrac OneFile; Ingenta; JCR Social Sciences Edition; Leisure, Recreation, and Tourism Abstracts; Online Computer Library Center FirstSearch Electronic Collections Online; Periodical Abstracts Library; Periodical Abstracts Research II; ProQuest 5000; ProQuest Social Science Plus Text; PsycINFO; Psychological Abstracts; Psychology and Behavioral Sciences Collection; Pollution Abstracts; Safety Science & Risk Abstracts; Sage Family Studies Abstracts; Social Services Abstracts; Social Science Citation Index; Sociological Abstracts; Student Resource Center Bronze; Student Resource Center Gold; Student Resource Center Silver; Violence & Abuse Abstracts; Wilson OmniFile V; Wilson Social Sciences, Abstracts; and Wilson Social Sciences Index*

Access to this journal is available free online within institutions in the developing world through the AGORA initiative with the FAO, the HINARI initiative with the WHO and the OARE initiative with UNEP. For information, visit www.aginternetwork.org, www.healthinternetwork.org, www.oaresciences.org

Child Development accepts articles for Open Access publication. Please visit http://olabout.wiley.com/WileyCDA/Section/id-406241.html for further information about OnlineOpen.

Wiley's Corporate Citizenship initiative seeks to address the environmental, social, economic, and ethical challenges faced in our business and which are important to our diverse stakeholder groups. Since launching the initiative, we have focused on sharing our content with those in need, enhancing community philanthropy, reducing our carbon impact, creating global guidelines and best practices for paper use, establishing a vendor code of ethics, and engaging our colleagues and other stakeholders in our efforts. Follow our progress at www.wiley.com/go/citizenship

Printed in the USA by The Sheridan Group

ISSN 0009-3920 (Print)
ISSN 1467-8624 (Online)

Published by Wiley for the Society for Research in Child Development

Child Development

Editor in Chief

Cynthia Garcia Coll
Carlos Albizu University

Associate Editors

Amy E. Booth, *Department of Psychology, Department of Communications and Science Disorders, University of Texas at Austin*
J. Gavin Bremner, *Psychology Department, Lancaster University*
Natasha J. Cabrera, *Department of Human Development and Quantitative Methodology, University of Maryland*
Carol McDonald Connor, *Psychology and Learning Sciences Institute, Arizona State University*
Michael Cunningham, *Department of Psychology, Tulane University*
David K. Dickinson, *Peabody College, Vanderbilt University*
Anne Fernald, *Department of Psychology, Stanford University*
Lisa A. Gennetian, *Institute for Human Development and Social Change, New York University*
Gustaf Gredebäck, *Department of Psychology, Uppsala University*
Amanda E. Guyer, *Department of Human Ecology and Center for Mind and Brain, University of California, Davis*
Marinus van IJzendoorn, *Centre for Child and Family Studies, Leiden University, Centre for Moral Socialization Studies, Erasmus University Rotterdam*
Bonnie J. Leadbeater, *Department of Psychology, University of Victoria, B.C. Canada*
Nonie K. Lesaux, *Harvard Graduate School of Education, Harvard University*
Katherine A. Magnuson, *School of Social Work, University of Wisconsin, Madison*
Tina Malti, *Department of Psychology, University of Toronto*
Diane Poulin-Dubois, *Department of Psychology, Concordia University*
Nirmala Rao, *Faculty of Education, The University of Hong Kong*
Glenn I. Roisman, *Institute of Child Development, University of Minnesota*
Megan Saylor, *Psychology and Human Development, Vanderbilt University*
Carlo Schuengel, *Department of Clinical Child and Family Studies, VU University Amsterdam*
Virginia Slaughter, *School of Psychology, University of Queensland*
Charles M. Super, *Department of Human Development & Family Studies, University of Connecticut*
Jennifer Brown Urban, *Department of Family and Child Studies, Montclair State University*
Liqi Zhu, *Institute of Psychology, Chinese Academy of Sciences*

Managing Editor

Adam Martin

Editorial Assistant

Stephanie Custer

Proofer

Kristin Hadfield

Consulting Editors

Jennifer Ablow
Emma K. Adam
Karen Adolph
Nazan Aksan
Joseph Allen
Christa Anderson
Dorit Aram
William F. Arsenio
Marian Bakermans-Kranenburg
Andreas Beelmann
Jay Belsky
Annie Bernier
Johanna Bick
Clancy Blair
Natasha K. Bowen
Julie Bowker
Amanda Brandone
Lee Branum-Martin
Keith Burt
Brad Bushman
Noel Card
Gustavo Carlo
Eva E. Chen
Bonnie Chow
Rebekah Coley
Katherine Conger
Lucy Cragg
Moritz Daum
Maayan Davidov
Mary Dozier
Dale Farran

Pasco Fearon
Ruth Feldman
Chris Fraley
Luciano Gasser
Roberta Golinkoff
Artin Goncu
Michaela Gummerum
David Haley
Carmel Houston-Price
Scott Johnson
Stephanie Jones
Ben Kenward
Maryam Kia-Keating
Ariel Kanfo
Jordy Kaufman
Ryan Kelly
Roger Kobak
Tobias Krettenauer
Kristin Lagattuta
Jonathan Lane
Amy Lederberg
Virginia Marchman
Catherine McBride-Chang
Kristina McDonald
Judi Mesman
Kathryn Monahan
Marlene Moretti
Frederick Morrison
Diana Murray-Close
Velma Murry
Nora Newcombe
Drew Nesdale
Mark Nielsen
Gil Noam
Lisa Oakes
David Oppenheim
Jamie Ostrov
Koraly Perez-Edgar
Beth Phillips
Eva Pomerantz
Douglas Powell
Vincent M. Reid
Cecil Reynolds
Marjorie Rhodes
Carolien Rieffe
Phil K. Rodkin
Martin Ruck
Christina Salmivalli
Chris Schatschneider
Brenda Schick
Carlo Schuengel
Monique Sénéchal
Patrick Shafto
Stefanie Sinno
Judi Smetana
Catherine Snow
Arlette Streri
Rebecca Treiman
Karine Verschueren
Henry Wellman
Rebecca White
Tuppet Yates

Child Development, September/October 2015, Volume 86, Number 5, Pages 1311–1671

Contents

EMPIRICAL ARTICLES

Individual Differences in Childhood Behavior Disorders Associated With Epigenetic Modulation of the Cortisol Receptor Gene
Mark R. Dadds, Caroline Moul, David J. Hawes, Antonio Mendoza Diaz, and John Brennan 1311

Attention to Faces Expressing Negative Emotion at 7 Months Predicts Attachment Security at 14 Months
Mikko J. Peltola, Linda Forssman, Kaija Puura, Marinus H. van IJzendoorn, and Jukka M. Leppänen 1321

Temperament and Interparental Conflict: The Role of Negative Emotionality in Predicting Child Behavioral Problems
Rochelle F. Hentges, Patrick T. Davies, and Dante Cicchetti 1333

24-Month-Old Children With Larger Oral Vocabularies Display Greater Academic and Behavioral Functioning at Kindergarten Entry
Paul L. Morgan, George Farkas, Marianne M. Hillemeier, Carol Scheffner Hammer, and Steve Maczuga 1351

Flexibility in Bilingual Infants' Word Learning
Katharine Graf Estes and Jessica F. Hay 1371

Prelinguistic Relational Concepts: Investigating Analogical Processing in Infants
Alissa L. Ferry, Susan J. Hespos, and Dedre Gentner 1386

Parents' Talk About Letters With Their Young Children
Rebecca Treiman, John Schmidt, Kristina Decker, Sarah Robins, Susan C. Levine, and Özlem E. Demir 1406

Cross-Linguistic Lexical, Grammatical, and Discourse Performance on Oral Narrative Retells Among Young Spanish Speakers
Audrey Lucero 1419

Knowing When to Be "Rational": Flexible Economic Decision Making and Executive Function in Preschool Children
Wendy S. C. Lee and Stephanie M. Carlson 1434

Hearing Where the Eyes See: Children Use an Irrelevant Visual Cue When Localizing Sounds
Karin Petrini, Peter R. Jones, Louise Smith, and Marko Nardini 1449

Episodic Future Thinking Is Related to Impulsive Decision Making in Healthy Adolescents
Uli Bromberg, Antonius Wiehler, and Jan Peters 1458

Early Nonparental Care and Social Behavior in Elementary School: Support for a Social Group Adaptation Hypothesis
Jean-Baptiste Pingault, Richard E. Tremblay, Frank Vitaro, Christa Japel, Michel Boivin, and Sylvana M. Côté 1469

Relationship Reciprocation Modulates Resource Allocation in Adolescent Social Networks: Developmental Effects
Stephanie Burnett Heyes, Yeou-Rong Jih, Per Block, Chii-Fen Hiu, Emily A. Holmes, and Jennifer Y. F. Lau 1489

On the Meanings of Self-Regulation:
Jeremy T. Burman, Christopher D. Green, and Stuart Shanker 1507

Expectations About Ethnic Peer Group Inclusivity: The Role of Shared Interests, Group Norms, and Stereotypes
Aline Hitti and Melanie Killen 1522

The Unique and Shared Genetic and Environmental Contributions to Fear, Anger, and Sadness in Childhood
Sierra Clifford, Kathryn Lemery-Chalfant, and H. Hill Goldsmith 1538

Reciprocal Relations Between Student–Teacher Relationship and Children's Behavioral Problems: Moderation by Child-Care Group Size
Věra Skalická, Jay Belsky, Frode Stenseng, and Lars Wichstrøm 1557

A Developmental Analysis of Caregiving Modalities Across Infancy in 38 Low- and Middle-Income Countries
Marc H. Bornstein, Diane L. Putnick, Jennifer E. Lansford, Kirby Deater-Deckard, and Robert H. Bradley 1571

Early Parenting and the Development of Externalizing Behavior Problems: Longitudinal Mediation Through Children's Executive Function
Michael J. Sulik, Clancy Blair, Roger Mills-Koonce, Daniel Berry, Mark Greenberg, and Mark Greenberg 1588

The Sexualized Girl: A Within-Gender Stereotype Among Elementary School Children
Ellen A. Stone, Christia Spears Brown, and Jennifer A. Jewell 1604

Distal Communication by Chimpanzees (Pan troglodytes): Evidence for Common Ground?
David A. Leavens, Lisa A. Reamer, Mary Catherine Mareno, Jamie L. Russell, Daniel Wilson, Steven J. Schapiro, and William D. Hopkins 1623

Quality Counts: Developmental Shifts in Associations Between Romantic Relationship Qualities and Psychosocial Adjustment
Charlene Collibee and Wyndol Furman 1639

Elementary School Children's Reasoning About Social Class: A Mixed-Methods Study
Rashmita S. Mistry, Christia S. Brown, Elizabeth S. White, Kirby A. Chow, and Cari Gillen-O'Neel 1653

OTHER

Manuscripts Accepted

For the complete list of Manuscripts Accepted for Publication, please visit www.srcd.org and navigate to ***Child Development*** *in the Publications section.*

Child Development, September/October 2015, Volume 86, Number 5, Pages 1311–1320

Individual Differences in Childhood Behavior Disorders Associated With Epigenetic Modulation of the Cortisol Receptor Gene

Mark R. Dadds, Caroline Moul, David J. Hawes, Antonio Mendoza Diaz, and John Brennan
The University of New South Wales

Behavioral/emotional difficulties in children are the first sign of mental health problems. These problems are however, heterogeneous. A domain that may identify homogenous subgroups is hypothalamic–pituitary–adrenal function. This study tested whether epigenetic regulation of glucocorticoid receptor gene could explain the co-occurrence of anxiety problems in children with behavior problems. Four- to 16-year-old clinically referred children (N = 241) were measured for psychiatric symptoms, methylation of target CpG sites in blood or saliva, and morning cortisol levels in those who gave blood. Increased methylation of promoter 1F CpG sites was associated with higher vulnerability to co-occurring internalizing symptoms and morning cortisol. The results support increasing optimism that epigenetic regulation of key neuroendocrine systems might help explain hitherto unfathomable individual differences in susceptibility to psychiatric symptom profiles.

There is a critical need to improve our understanding of the first signs and developmental trajectories of mental health problems. Although psychiatric diagnoses can be reliably differentiated by their presenting symptoms (e.g., attention deficit hyperactivity disorder [ADHD]), existing diagnostic systems are unable to specify discrete categories with precise etiologies, prognoses, and treatment needs. Consensus is emerging on the need for neurodevelopmental models that consider "disorder" at the level of interrelations among neural function, environment, and behavior rather than the notion of discrete diagnostic typologies (Insel et al., 2010; Kendler, Zachar, & Craver, 2011).

One area where this need is paramount is early-onset behavioral or conduct problems (CPs) in children. These are the most common child mental health condition and the most common precursor to all forms of adult mental health and social problems (Copeland, Shanahan, Costello, & Angold, 2009; Kim-Cohen et al., 2003). While CP children show commonalities of impulsive, aggressive, and rule-breaking behavior, existing diagnostic categories have little predictive power in terms of prognosis and treatment response. Similarly, the search for environmental conditions that differentiate more homogeneous groupings has yielded little fruit. That is, the major environmental risk factors identified for criminality, mental health disorders, substance use, and even physical health problems, reveal more that is common and generic (e.g., child abuse, poverty, family disruption, trauma) than specific (e.g., World Health Organization, 2008).

In contrast, research looking at how environmental conditions trigger, inhibit, and interact with specific biological vulnerabilities has the potential to yield more information (Belsky & Pluess, 2013; Pluess & Belsky, 2012). For example, variations in prognosis and treatment response can be predicted by variations in critical neurodevelopmental systems associated with motivation/emotionality, social attention/cognition, and inhibition/impulsivity (Matthys, Vanderschuren, Schutter, & Lochman, 2012; Schechter, Brennan, Cunningham, Foster, & Whitmore, 2012). CPs are commonly but not always, associated with high emotional lability that specifies risk of developing anxiety, depression, and substance-use problems (Copeland et al., 2009). Children with high emotional lability typically show "hot-tempered" or reactive aggression, their problems develop largely through poor parenting and emotional dysregulation (Viding et al., 2010), and they respond well to evidence-based treatments (Brestan & Eyberg, 1998). In contrast, children characterized by low emotionality (or callous-unemo-

---

The authors acknowledge the support of ARC Linkage Grants LP100200150 and NHMRC Project Grant 568667 to the first author, and the partnership with Royal Far West in recruiting and supporting the participant families. The authors have no financial disclosures to make.

Correspondence concerning this article should be addressed to Mark Dadds, School of Psychology, University of New South Wales, Sydney, NSW, Australia 2052. Electronic mail may be sent to m.dadds@unsw.edu.au.

© 2015 The Authors
Child Development © 2015 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2015/8605-0001
DOI: 10.1111/cdev.12391

tional traits) have relatively higher genetic influence and respond relatively poorly to treatment (Hawes, Price, & Dadds, 2014). High levels of concurrent anxiety in CP children, although conferring risk for adult mental health problems, may protect against extreme antisocial outcomes (Washburn et al., 2007), and predict positive responses to psychological interventions for ADHD (March et al., 2000). Thus, the concurrence of high emotionality is important for understanding more homogeneous trajectories and treatment needs of children with early-onset mental health problems.

An important target in this regard is the hypothalamic–pituitary–adrenal (HPA) axis, the neuroendocrine system that mediates the body's reactions to stress through regulation of diverse processes from digestion to mood to energy storage and expenditure, as well as emotional and cognitive function. Disturbances in the HPA axis are thought to play an important role in both the development of conduct (or externalizing) problems and internalizing (anxiety, depression) disorders. For externalizing disorders, research has largely focused on direct associations between symptom levels and circulating cortisol and has often produced divergent findings (Hawes, Brennan, & Dadds, 2009). For example, aggressive/antisocial behavior is often characterized as being associated with down regulation of the HPA and low cortisol levels, whereas many other studies find aggressive antisocial children show elevated levels (see Hawes et al., 2009).

We hypothesize that some of this inconsistency might be associated with the co-occurrence of internalizing problems in these children. Research shows that HPA axis dysfunction plays a key role in the neurobiology of mood and anxiety disorders (including posttraumatic stress); many antidepressants have a therapeutic impact via regulation of HPA axis function. Chronic and uncontrollable stress leads to chronic overactivation of the HPA axis and high levels of circulating glucocorticoids (cortisol in humans) associated with increased risk for anxiety, depression, and a range of physical health problems (Lupien, McEwen, Gunnar, & Heim, 2009). Regulation of the HPA axis involves feedback loops involving central nervous system glucocorticoid receptors (GR) that influence corticotropin-releasing factor and thus ACTH (Adrenocorticotropic hormone) from the anterior pituitary. Animal and human research has shown that positive early parenting experiences alter GR distribution and function, particularly in the hippocampal region. GR distribution and function affect the probability of adaptive responses to acute stress through HPA axis activation (Meaney & Ferguson-Smith, 2010). Following major early stress such as child abuse and neglect, however, there is low transcription of functional receptor networks impeding negative feedback and resulting in chronically high levels of circulating cortisol. A corollary is increased lifelong risk for physical and mental health disorders, especially those concerning anxiety and depression (Meaney & Ferguson-Smith, 2010).

The transcription of neural GRs is controlled by the cortisol receptor NR3C1 gene and its influence is regulated in part by epigenetic processes such as methylation, whereby gene transcription is altered due to the specific binding of methyl molecules to "islands" of CpG sites. CpG sites are regions of DNA where a cytosine nucleotide occurs next to a guanine nucleotide in the linear sequence of bases along its length, linked together by phosphate binding. In mammals, 70%–80% of CpG cytosines are methylated, and methylation can be triggered by a variety of biological and environmental stimuli (Meaney & Ferguson-Smith, 2010; Szyf & Bick, 2013). Increased methylation generally leads to reduced mRNA signaling and the resulting protein product. Important environmental processes such as quality of parenting and abuse history can lead to methylation of genes critical for neurodevelopment (van IJzendoorn, Bakermans-Kranenburg, & Ebstein, 2011; Szyf, 2013). Furthermore, some DNA methylation states appear to be reversible (Dulac, 2010).

Individual differences in methylation of specific CpG sites in the NR3C1 promoter region 1F (Turner et al., 2010) have been shown to be both sensitive to the quality of early environments and predictive of mental health outcomes. That is, adverse early environments lead to higher methylation, less GR transcription and functionality, and risk for anxiety and depression (Meaney & Szyf, 2005). Table 1 summarizes key existing research into methylation of the CpG island on the IF promoter region described by McGowan et al. (2009). A number of studies show that increased methylation of the region spanning chr5:142783689–142783638 of NR3C1 promoter 1F CpG sites is associated with less functional transcription of GR receptors, higher circulating cortisol levels, exposure to early adverse environments, and specific vulnerability to internalizing symptoms of anxiety and depression.

Research into epigenetic regulation of stress function is one of the most comprehensive and precise models yet developed in psychiatry, in that it spans and integrates data about precise environmental processes (maternal care) through epigenetic modifications to neural function, to psychiatric phenotypes

Table 1

CpG Sites Assayed in the Cortisol Receptor NR3C1 Gene, chr5 Promoter 1F for the Current Study and Corresponding to McGowan et al. (2009)

McGowan et al. (2009)		Current study	Significant methylation effects
CpG 1	chr5:142783884	CpG 1	
CpG 2	chr5:142783874	CpG 2	
CpG 3	chr5:142783870		
CpG 4	chr5:142783864		
CpG 5	chr5:142783860	CpG 3	
CpG 6	chr5:142783858		
CpG 7	chr5:142783854		
CpG 8	chr5:142783849	CpG 4	Maternal emotional state during pregnancy[a]
CpG 9	chr5:142783844		
CpG 10	chr5:142783838	CpG 5	
CpG 11	chr5:142783832		
CpG 12	chr5:142783822	CpG 6	Functional[b]
CpG 13	chr5:142783810	CpG 7	Functional[b]
CpG 23	chr5:142783745	CpG 8	
CpG 24	chr5:142783743		
CpG 29	chr5:142783703	CpG 9	
CpG 30	chr5:142783689	CpG 10	Functional and abuse related[b]
CpG 31	chr5:142783686		Functional[b]
CpG 32	chr5:142783679		Functional and abuse related[b]
CpG 35	chr5:142783640	CpG 11	Maternal depression[c] Early parent death[d] Maternal emotional state during pregnancy[a]
CpG 36	chr5:142783638		Maternal depression[c]
CpG 37	chr5:142783628	CpG 12	Maternal depression; infant stress cortisol[c] Abuse related[e]
CpG 38	chr5:142783622		

[a] Hompes et al., 2013; [b] McGowan et al., 2009; [c] Oberlander, Weinberg, Papsdorf, Grunau, Misri, and Devlin 2008; [d] Melas et al., 2013; [e] Romens McDonald, Svaren, and Pollak 2014.

(Szyf & Bick, 2013). While first developed on animals, there is growing evidence that these methylation effects apply equally well to humans (McGowan et al., 2009), including newborns (Bromer, Marsit, Armstrong, Padbury, & Lester, 2013; Hompes et al., 2013; Oberlander et al., 2008). This raises the question of whether differential methylation of NR3C1 may help explain the heterogeneity of early-onset mental health problems in children in terms of their diagnostic profile and cortisol function.

Thus, the aim of this study was to conduct the first test of the role of NR3C1 methylation in emerging psychopathology in children. Specifically, we tested whether individual differences in epigenetic regulation of the HPA function can explain individual differences in the symptom profiles of children with early-onset behavior problems. Given the emerging evidence of the specific relation of epigenetic regulation of HPA function to anxiety and depression (internalizing disorders), we hypothesized that methylation levels would be associated with severity of children's comorbid internalizing, but not externalizing, disorders. Specifically, methylation of CpG sites previously identified as reflecting vulnerability to internalizing disorders (GR gene NR3C1 promoter 1F CpG sites (based on McGowan et al., 2009) would be positively correlated with internalizing symptoms (not externalizing symptoms).

Method

Participants

Ethics approval was from the University of New South Wales (UNSW). Participants were two samples of children, totaling 241 children (79% male), referred for the treatment of behavioral/emotional problems to the Child Behaviour Research Clinic at the UNSW, and Royal Far West, Sydney. All children: (a) met formal criteria for full or features of a *Diagnostic and Statistical Manual of Mental Disorders*, 4th ed. (*DSM–IV*) disruptive behavior of oppositional-defiant (ODD) or CD, and typically had concurrent anxiety or mood, or developmental disorders (severity rating 0–6 where > 3 = frank diagnosis; American Psychiatric Association, 1994); (b) were aged from 4 to 16 years ($M = 8.76$, $SD = 3.21$); (c) had functional English; (d) had no major neurological/physical illness; and (e) had $IQ > 70$. Ethnic ancestry was measured by having parents nominate the ancestry of each grandparent as either Anglo-European (Caucasian), Asian/Indian, Middle Eastern/North African, Indigenous, Oceanic/Pacific, and Other. Children were then categorized as either Caucasian (three or more grandparents known and rated as Caucasian), Other, or Not Known. The sample was largely Caucasian with small percentages of Asian, Middle Eastern, and Indigenous heritage (see Table 2). In terms of number of diagnoses, 38% had one diagnosis, 49% had two, and 13% had three or more comorbid conditions. For primary diagnoses, 66% met criteria for CP (ODD or CD), 26% ADHD, 6% anxiety or depression, and 5% for features of autism spectrum disorder (ASD). For secondary diagnoses, 26% met criteria for CP, 17% ADHD, 14% anxiety or depression, and 3% features of ASD.

Table 2
Demographic Data and Psychological Adjustment of the Two Samples

	Blood ($N = 91$)	Saliva ($N = 156$)
Age	8.66 (3.09)	8.03 (3.21)
% Male[a]	86.80	74.40
% Caucasian	79.30	77.00
SEIFA rank[a]	7.09 (2.56)	7.53 (2.72)
Total diagnostic severity	2.23 (0.64)	2.47 (0.87)
Severity externalizing symptoms	4.13 (0.79)	4.05 (0.81)
Severity ADHD	2.14 (2.00)	1.70 (2.01)
Severity internalizing symptoms[a]	0.33 (0.96)	0.87 (1.58)

Note. SEIFA = Socioeconomic Indexes for Areas (Australian Bureau of Statistics); ADHD = attention deficit hyperactivity disorder.

[a] Means are different at $p < .05$.

Diagnostic Procedures and Measures

Diagnoses were made using *DSM–IV* criteria by the assessing psychiatrist/psychologist using the Diagnostic Interview Schedule for Children, Adolescents, and Parents (Holland & Dadds, 1997) with parents, and the child for those older than 8 years. Diagnoses were checked by having a second diagnostic team make an independent diagnosis. Kappa agreements on primary and secondary diagnoses were always > 0.7. Final measures of diagnostic severity used for analyses were Severity of Externalizing (ODD and CD) and Severity of Internalizing (any Anxiety and Depression diagnoses), both measured on 0–6 point scales. Child symptomology was also measured using the Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997). The measure has good psychometric properties and converges well with other indices of child mental health including independent *DSM* diagnoses. Given our focus on the broad externalizing and internalizing domains, we used SDQ Total Score, and the CP and Anxiety/Depression subscales. Reports were available for all mothers and subsets of fathers and teachers. Convergence with mothers' scores is strong; for example, Total Score—father $r = .662, p < .05$; teacher $r = .415, p < .05$.

Biological Measures

Participants gave saliva or blood at local pathology collection clinics. Whole blood was collected in ethylene-diamine-tetra-acetate anticoagulant vials (BD Vacutainers, Becton, Dickinson & Co., Franklin Lakes, NJ) and DNA was recovered from whole blood by silica matrix purification (2.0 ml; NucleoSpin® Blood L, cat. no. 740954; Macherey-Nagel GmbH & Co. KG, Düren, Germany) according to manufacturer's instructions. Saliva was collected using Oragene collection and stabilization systems (Oragene DNA [OG-500]; DNA Genotek Inc., Kanata, Canada; DNA was recovered from stabilized saliva samples using the manual prepIT system according to manufacturer's instructions). DNA precipitates were allowed to resuspend for a minimum of 48 hr before quantification by fluorimetry (QuantiFluor™ dsDNA System; Promega Corporation (Madison, WI) in conjunction with a Gemini™ Spectramax XPS fluorescence microplate reader (Molecular Devices, LLC, Sunnyvale, CA). DNA stocks were adjusted to a working concentration of between 10 and 50 ng/µl for subsequent genotyping. DNA extraction and genotyping were performed by the Australian Genome Research Facility (AGRF) Ltd; DNA extraction rates were > 95%.

Samples were genotyped using iPLEX Gold™ primer extension followed by mass spectrometry analysis on the Sequenom MassARRAY system (Sequenom, San Diego, CA, USA) by the AGRF (http://www.agrf.org.au/). GR methylation levels were quantified using the EpiTYPER assay on the Sequenom MassARRAY system at the AGRF. Percent methylation was assayed for the CpG island in the GR gene NR3C1 promoter 1F (similar to rat exon 17, which reveals an effect of maternal behavior on cytosine methylation and expression). The island was chosen based on McGowan et al. (2009): Left primer: aggaagagagTTTTTGAAGTTTTTTTAGAGG G, Right primer: cagtaatacgactcactatagggagaaggctA ATTTCTCCAATTTCTTTTCTC; Amplicon 5′ start: chr5:142783906, Amplicon 3′ end: chr5:142783586. Valid signals were achieved for the 12 CpG sites shown in Table 1. Reliability of the methylation assays was checked for each of these CpG sites. Epitect control DNA samples (Qiagen, Valencia, CA) known to be fully methylated and fully unmethylated were assessed in parallel to target CpG sites to determine the upper and lower limits of detection for each assay. Mean methylation levels were measured at 93% and 5.3% for fully methylated and fully unmethylated DNA, respectively.

Children were asked to give blood as the default but saliva collection was given as a secondary option and 61.3% chose this. There is some evidence that methylation changes are systemwide and can be indexed reliably across tissue systems (Auta et al., 2013; Rosas et al., 2001; Szyf & Bick, 2013), especially blood and saliva (Rosas et al., 2001); however, convergence across blood and saliva for GR methylation is unknown, so we tested hypotheses within the blood ($N = 90$) versus saliva ($N = 151$) cohorts separately.

Morning plasma cortisol levels were assayed for the cohort who gave blood samples. These were collected at pathology collection centers between 8:00 a.m. and 10:00 a.m. Children attended with at least one caregiver. Upon presentation at the center, the nurse-phlebotomist escorted the child (and parent) to a calm environment to sit quietly and rest for 10 min. Blood was then extracted from the upper arm and the child was rewarded with stickers and a certificate. Nurses and parents dealt with mild distress by reassuring and calming the child; however, any continuing distress or refusal was taken as revocation of consent and the procedure was abandoned. Samples remained at room temperature for 30 min before spinning at 3,000 rotations per minute for 5–10 min. Serum was stored at −40°C until analysis. Cortisol levels were determined at Sydney South West Pathology Services using Chemiluminescent Enzyme Immunoassay (Siemens Immulite 2000; Siemens Medical Solutions Diagnostics, Bayswater, Australia; Limit of detection of 5.5 nmol/L, reference range of 200–600 nmol/L). Intra-assay and interassay variations were 5.2% and 6.8%, respectively.

Statistical Design

We first checked the cohort for levels of mental health problems compared to a representative community sample, differences in methylation and cortisol levels according to ethnic ancestry, gender, age and for methylation levels, and cohorts who gave blood versus saliva. We then examined convergence among methylation levels at the individual CpG sites. Previous research has variously looked at methylation at individual CpG sites as well as mean methylation levels across sites (Romens, McDonald, Svaren, & Pollak, 2014; Szyf & Bick, 2013). Little is known about the convergence of sites within the CpG island of NR3C1. We analysed methylation across sites on the assumption that there would be a common and unique variance in methylation levels across individual CpG sites. To index common variance in methylation levels, we estimated the number of unique components in the methylation data using Velicer's Minimum Average Partial test (O'Connor, 2000). This procedure compares eigenvalues derived from principal component analysis of the data to those obtained by chance using Monte Carlo data sets. This suggested one component was present in the methylation levels. Table 3 shows component loadings for methylation levels at each CpG site on to the common methylation index.

Table 3
Component Matrix Showing Factor Loadings > 0.3 for Methylation Levels at Each CpG Site

Methylation site	Component loading
CpG1	0.51
CpG2	—
CpG3	0.30
CpG4	—
CpG5	—
CpG6	0.65
CpG7	—
CpG8	0.46
CpG9	0.47
CpG10	0.54
CpG11	—
CpG12	0.64

Note. Extraction: principal component analysis.

The hypothesis that levels of methylation would predict levels of internalizing symptoms were tested using a series of regression analyses within blood and saliva tissue cohorts separately. Age and ethnic ancestry were used as covariate predictors in all analyses; gender and socioeconomic status (Socio-Economic Indexes for Areas [SEIFA]) were tested but failed to show significance in any analysis and so were omitted. First, to test associations for the overall index of methylation, age, ethnic ancestry (dummy coded Caucasian vs. non-Caucasian), and severity of externalizing and internalizing disorders, were used as predictors of methylation levels in a stepwise regression analysis using SPSS v.22 (IBM Corporation, New York, NY). Second, we used methylation levels of individual CpG sites, age, and ethnic ancestry in a stepwise regression to predict levels of internalizing and externalizing psychopathology. To control for multiple comparisons for the individual CpG sites, we looked for replication of associations across tissue sample groups as well as significance at Bonferroni adjusted levels. To assess associations between methylation and morning cortisol levels, we again used the overall index of methylation and then methylation levels at each CpG site. Cortisol levels, measured in the blood cohort only, were the dependent variable and age, time at which the morning sample was taken, ethnic ancestry, and methylation levels were used as predictors.

Results

Table 4 shows means and standard deviations on the subscales of the SDQ for the current cohort

Table 4
Means and Standard Deviations *on Total Scores and Subscales of the Strengths and Difficulties Questionnaire for the Current Sample Compared to an Australian Normative Sample (Mellor, 2005)*

	Emotional	Conduct problems	Hyperactivity	Peer problems	Prosocial	Total difficulties
Possible range	0–10	0–10	0–10	0–10	0–10	0–40
Australian norms						
M	2.1	1.5	3.1	1.6	8.3	8.18
SD	2.0	1.6	2.4	1.9	1.7	6.06
Range	0–8	0–6	0–10	0–8	0–10	0–26
Current sample						
M	3.90	4.52	6.63	3.47	5.86	18.53
SD	2.68	2.35	2.70	2.38	2.32	6.94
Range	0–10	0–10	0–10	0–10	0–10	0–37

compared to normative data from Mellor (2005). As expected, this sample shows significant elevations across both domains of externalizing and internalizing problems. Table 2 shows demographic and diagnostic information for the two cohorts. We checked whether there were cohort differences in cortisol and methylation levels by ethnic ancestry, gender, age, and for methylation levels, those who gave blood versus saliva. There were no differences in cortisol levels for gender or age group, although older children tended to give blood later in the morning, $F(1, 74) = 5.49$, $p < .05$. Age and time of sample being taken were used as covariates in all analyses involving cortisol levels. Cortisol levels also differed by ethnic ancestry, $F(2, 72) = 4.77$, $p < .05$, with non-Caucasian group ($M = 308.37$, $SD = 154.59$) showing higher levels than the Caucasian group ($M = 230.54$, $SD = 90.21$). There were no overall differences in methylation levels by ethnic ancestry, age group, or gender.

We checked cohort differences for those who gave saliva versus blood. There were no differences by ethnic ancestry. Those who gave saliva were younger ($M = 8.38$, $SD = 3.21$ vs. $M = 9.35$, $SD = 3.11$), $F(1, 344) = 7.597$, $p = .006$; more likely to be female (76% vs. 57% within males vs. females); higher socioeconomic status (SEIFA; $M = 7.28$, $SD = 2.72$ vs. $M = 6.54$, $SD = 2.41$), $F(1, 295) = 5.76$, $p = .018$, less severe on ADHD diagnostic severity ($M = 1.80$; $SD = 1.98$ vs. $M = 2.33$, $SD = 1.95$), $F(1, 335) = 5.66$, $p = .018$, and more severe on anxiety disorders ($M = 1.10$, $SD = 1.75$ vs. $M = .51$, $SD = 1.28$), $F(1, 336) = 11.11$, $p = .001$. When these variables were entered as simultaneous predictors of blood versus saliva cohorts in a logistic regression in SPSS, the significant predictors were gender (Wald = 10.39, $p = .001$) and anxiety (Wald = 9.79, $p = .002$), whereby those opting to give saliva were more likely to be female and highly anxious.

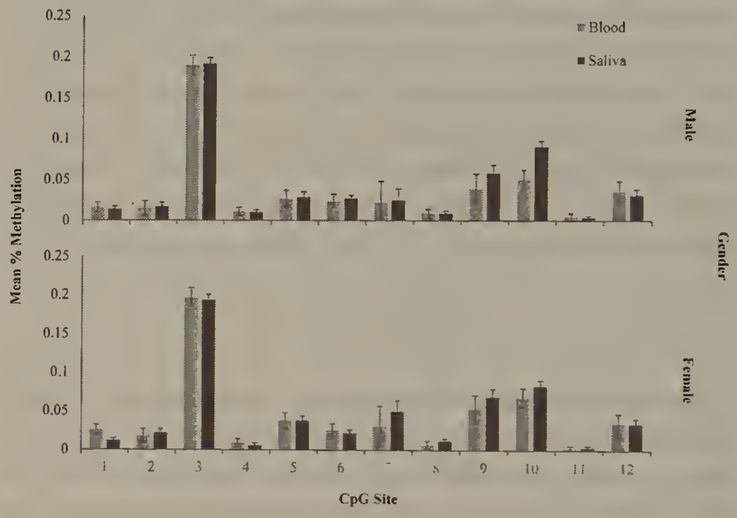


Figure 1. Mean methylation levels across the 12 NR3C1 CpG sites split by tissue type and gender. Error bars represent 1 *SE*.

Unstandardized percent methylation levels for each of the 12 CpG sites are shown in Figure 1. Patterns of methylation are remarkably similar across blood and saliva. Analyses of variance comparing methylation levels for the 12 sites by tissue source showed that there were significant differences only for Sites 9 and 10 where higher methylation levels were found for saliva versus blood samples:CpG 9 ($M = 0.057$, $SD = 0.068$ vs. $M = 0.038$, $SD = 0.043$), $F(1, 345) = 8.36$, $p = .004$, and CpG 10 ($M = 0.091$, $SD = 0.045$ vs. $M = 0.053$, $SD = 0.033$), $F(1, 345) = 69.96$, $p < .001$. CpG site 10 has previously been found by McGowan et al., 2009 to be both functional and abuse related and features in our hypotheses about heightened methylation and risk for anxiety disorders.

Table 5

Results of Stepwise Regressions Predicting Total Diagnostic Anxiety/Depression, Conduct Problems, and Morning Cortisol Levels, From Ethnic Ancestry, Age, and % Methylation for Specific NCR1 CpG Sites

	Blood (n = 91)				Saliva (n = 156)			
	Predictor	Beta weight	t	p	Predictor	Beta weight	t	p
Internalizing symptoms	CpG 1	.21	2.03	.041				
	CpG 10	.33	3.30	.003^{a}	CpG 10	.17	2.11	.03
Externalizing symptoms	Caucasian	.21	2.02	.047	Age	.25	3.26	.001^{a}
					CpG 3	.19	2.40	.01
					CpG 4	.26	3.18	.001^{a}
Morning cortisol	Time taken	−.25	2.43	.018				
	CpG 9	.44	4.21	.0001^{a}				

Note. Time taken = time of day at which blood was taken for cortisol assay.
^{a}Significant at Bonferroni adjusted levels.

Common Index of Methylation Levels

First we tested associations with the overall index of methylation levels by regressing age, ethnic ancestry, and severity of externalizing and internalizing disorders and age, onto methylation levels. For blood, the regression equation was significant, $F(1, 85) = 2.39$, $p < .05$, ethnic ancestry ($\beta = -.23$; Caucasian associated with lower methylation levels) and levels of internalizing disorders ($\beta = .23$) being the significant predictors. This significant prediction of methylation levels was replicated in the saliva sample for internalizing disorders ($\beta = .175$), but not ethnic ancestry ($\beta = -.009$). Thus, higher levels of internalizing disorders were uniquely associated with higher levels of methylation across the CpG sites in both tissue sources. The regression predicting morning cortisol levels from ethnic ancestry, age, time of sample being taken, and common methylation levels was significant, $F(4, 67) = 3.89$, $p < .01$. Three predictors were significant in that being non-Caucasian ($\beta = -.21$), older ($\beta = .22$), and having higher methylation levels ($\beta = .23$), were all associated with higher morning cortisol levels.

Methylation at the Individual CpG Sites

Associations were determined by regressing age, ethnic ancestry, and methylation levels for the 12 CpG sites onto levels of internalizing and externalizing disorders. Statistical reliability was estimated by examining replication across cohorts and within cohorts using Bonferroni adjusted significance of $p < .004$ (.05 ÷ 12 CpG sites). The results are shown in Table 5. As hypothesized, increased methylation of specific CpG sites was associated with severity of internalizing symptoms. Consistent with McGowan et al. (2009), increased methylation of CpG site 10 was associated with higher internalizing symptoms for both tissue cohorts. CpG Sites 3 and 4 predicted externalizing severity; however, these associations did not replicate across tissues. For the blood cohort, higher morning cortisol levels were independently associated with earlier time of blood extraction and increased methylation of CpG 9.

Discussion

Much has recently been made of the potential role of epigenetic mechanisms to explaining environmental–biological interactions in the development of mental health problems (e.g., Szyf & Bick, 2013). A recent and well-characterized model of epigenetic variation shows that individual differences in vulnerability can occur via epigenetic regulation of the GR gene. Thus, exposure to early environmental adversity produces increased methylation of GR promoter regions, which in turn leads to disrupted neural GR transcription and function and poorly regulated HPA axis function, long known to be associated with anxiety and depression (McGowan et al., 2011).

The current study conducted the first test of this methylation pattern in young children showing emerging psychopathology. The results showed that epigenetic regulation of GR can help explain individual differences in diagnostic or phenotypic profiles of children showing early-onset behavioral/emotional problems. Specifically, increased epigenetic methylation of the IF promoter region, especially a region spanning chr5: 142783689–142783638 of NR3C1 promoter 1F CpG sites (McGowan et al., 2009; sites CpG29–36) is associated with both higher circulating cortisol levels and increased vulnerability to anxiety and depression in young children.

As noted, early-onset behavioral problems are one of the earliest and most reliable signs of developing mental health and social adjustment problems. Children showing these problems are however, a heterogenous group and differences in their life paths have been difficult to understand and predict. The current data show that epigenetic regulation of the cortisol receptor system, corresponding to well-characterized animal models of stress and epigenetic regulation of HPA axis function (Meaney & Szyf, 2005), may in part explain the lifelong higher risk a substantial subset of these children show for disorders of anxiety and depression.

Similar to McGowan et al. (2009), we found that methylation of this CpG region chr5:142783689–142783638 was related to both psychological measures and HPA function. That is, increased methylation was related to higher levels of internalizing psychiatric symptoms and higher circulating cortisol. A previous study using newborn cord blood and salivary cortisol in the infants at 3 months showed links between maternal depression and the infant's methylation of this NR3C1 region and circulating cortisol levels (Oberlander et al., 2008). Our data confirm the link between methylation in this region and circulating cortisol in older children and show that it is associated with the emergence of anxiety and depression.

With respect to environmental adversity, we were unable to test the idea that methylation of these NR3C1 sites stems from exposure to early adversity, as all of our available measures of adversity were taken subsequent to the development of the child's psychopathology and not targeted at early infancy prior to the onset of the mental health problems. Thus, the adversity we indexed was likely to be a result of the child's problems. Our guiding model (Meaney & Szyf, 2005) specifies that early adversity occurs prior to, and leads to, epigenetic changes (methylation) that are then concurrent with higher circulating cortisol and vulnerability to anxiety and depression. It would be inappropriate to test the relation of epigenetic changes, cortisol, and psychopathology to adversity, if that adversity has arisen from the psychopathology.

Other aspects of the current data should be noted. Methylation studies are in their infancy and little is known about how to treat varying methylation levels across large numbers of CpG islands and sites. Some researchers take global or mean methylation levels (e.g., Herbstman et al., 2013). Our data indicate that this could be metrically unsound as we were unable to develop a sound index of global methylation that included all CpG sites. Furthermore, when we analyzed methylation at the level of individual CpG sites, the significant associations with internalizing disorders were replicated for one site only, a site that has previously been shown to have both functional significance and associations with internalizing disorders (McGowan et al., 2011). At the level of individual sites, however, this site was not associated with circulating cortisol levels; rather, the latter were associated with an adjacent CpG site. Until more is known about common and unique variance in methylation across CpG islands versus individual sites, this lack of convergence is difficult to interpret. Limitations in our measurement of cortisol could in part explain this.

Ideally, circulating cortisol should be measured multiple times across diurnal variation, controlling for time of awakening, and in response to rest versus a stress challenge rather than a single time point as done in this study. Furthermore, cortisol assays were only available for the subsample that gave blood, are thus underrepresented by highly anxious children, and are thereby likely to underestimate the size of the relation to methylation levels. As noted, our sample was measured at one time point after the development of symptoms and it is impossible to make causal or even temporal conclusions about the relation of these symptoms and exposure to environmental adversity.

Methylation was measured in both blood and saliva for subsamples who gave consent for one method over another. There is some evidence indicating that epigenetic changes are systemwide; however, the applicability of this to blood and saliva, and how these converge with neural transcription of the NR3C1 gene are unknown. The similar patterns of methylation across blood and saliva, and replication of the association of CpG site 10 to internalizing problems across tissues, are consistent with the idea that methylation occurs similarly across these tissue systems; however, more targeted research will be needed to check the veracity of this. As whole blood was used, we cannot specify which cellular subtypes showed the epigenetic features within the blood samples. Finally, our cohort was biased to males showing externalizing problems and spanned a considerable age range. No age or gender differences in methylation were found; however, we cannot be certain that there were no gender- or age-related differences in complex relation among methylation, adversity, and symptoms. Finally, all our analyses were conducted within a clinical sample. Although we tested directional hypotheses that higher methylation would be associated with higher cortisol and more psychopathology, we cannot assume that the methylation levels we observed are elevated compared to community nonclinical samples.

In summary, the current study shows that epigenetic mechanisms of regulation of the GR gene can help explain individual differences in diagnostic or phenotypic profiles of early-onset mental health problems. Consistent with several previous studies, most notably the seminal study by McGowan et al. (2009), we showed that increased epigenetic methylation of a region spanning chr5:142783689–142783638 of NR3C1 promoter 1F CpG sites is associated with both higher circulating cortisol levels and specific vulnerability to internalizing symptoms of anxiety and depression in young children with CP. High levels of concurrent anxiety in CP children are a mixed phenomenon; they confer extra risk for adult mental health problems but may also protect against extreme antisocial outcomes (Washburn et al., 2007) and predict positive responses to psychological interventions for ADHD (March et al., 2000). Thus, the concurrence of high emotionality is important for understanding more homogeneous trajectories and treatment needs of children with early-onset mental health problems. In terms of the methylation patterns, the site at chr5:142783679 may be particularly interesting as it was uniquely associated with internalizing disorders and has been previously shown to be both functional in the cortisol receptor system and predictive of risk for internalizing disorders (see Table 1). It has long been known that exposure to adverse early childrearing conditions is associated with increased risk for lifelong mental health problems; however, individual differences in the susceptibility to adversity and health outcomes have been difficult to explain. The current results support increasing evidence that epigenetic mechanisms have the potential to help explain complex environmental–biological interactions in the development of mental health problems.

References

American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders*, 4th ed.. Washington, DC: Author.

Auta, J., Smith, R. C., Dong, E., Tueting, P., Sershen, H., Boules, S., . . . Guidotti, A. (2013). DNA-methylation gene network dysregulation in peripheral blood lymphocytes of schizophrenia patients. *Schizophrenia Research, 150*, 312–318. doi:10.1016/j.schres.2013.07.030

Belsky, J., & Pluess, M. (2013). Beyond risk, resilience, and dysregulation: Phenotypic plasticity and human development. *Development and Psychopathology, 25*(4, Pt. 2), 1243–1261. doi:10.1017/S095457941300059X

Brestan, E. V., & Eyberg, S. M. (1998). Effective psychosocial treatments of conduct-disordered children and adolescents: 29 years, 82 studies, and 5,272 kids. *Journal of Clinical Child Psychology, 27*, 180–189. doi:10.1207/s15374424jccp2702_5

Bromer, C., Marsit, C. J., Armstrong, D. A., Padbury, J. F., & Lester, B. (2013). Genetic and epigenetic variation of the glucocorticoid receptor (NR3C1) in placenta and infant neurobehavior. *Developmental Psychobiology, 55*, 673–683. doi:10.1002/dev.21061

Copeland, W., Shanahan, L., Costello, E., & Angold, A. (2009). Childhood and adolescent psychiatric disorders as predictors of young adult disorders. *Archives of General Psychiatry, 66*, 764–772. doi:10.1001/archgenpsychiatry.2009.85

Dulac, C. (2010). Brain function and chromatin plasticity. *Nature, 465*, 728–735. doi:10.1038/nature09231

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 38*, 581–586. doi:10.1111/j.1469-7610.1997.tb01545.x

Hawes, D. J., Brennan, J., & Dadds, M. R. (2009). Cortisol, callous-unemotional traits, and pathways to antisocial behavior. *Current Opinion in Psychiatry, 22*, 357–362. doi:10.1097/YCO.0b013e32832bfa6d

Hawes, D. J., Price, M. J., & Dadds, M. R. (2014). Callous-unemotional traits and the treatment of conduct problems in childhood and adolescence: A comprehensive review. *Clinical Child and Family Psychology Review, 17*, 248–267. doi:10.1007/s10567-014-0167-1

Herbstman, J. B., Wang, S., Perera, F. P., Lederman, S. A., Vishnevetsky, J., Rundle, A. G., . . . Tang, D. (2013). Predictors and consequences of global DNA methylation in cord blood and at three years. *PLoS ONE, 8*, e72824. doi:10.1371/journal.pone.0072824

Holland, D., & Dadds, M. R. (1997). *The Diagnostic Interview Schedule for Children, Adolescents and Parents*. Brisbane, Australia: Griffith University.

Hompes, T., Izzi, B., Gellens, E., Morreels, M., Fieuws, S., Pexsters, A., . . . Claes, S. (2013). Investigating the influence of maternal cortisol and emotional state during pregnancy on the DNA methylation status of the glucocorticoid receptor gene (NR3C1) promoter region in cord blood. *Journal of Psychiatric Research, 47*, 880–891. doi:10.1016/j.jpsychires.2013.03.009

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., . . . Wang, P. (2010). Research Domain Criteria (RDoC): Toward a new classification framework for research on mental disorders. *American Journal of Psychiatry, 167*, 748–751. doi:10.1176/appi.ajp.2010.09091379

Kendler, K. S., Zachar, P., & Craver, C. (2011). What kinds of things are psychiatric disorders? *Psychological*

Medicine, 41, 1143–1150. doi:10.1017/S0033291710001844

Kim-Cohen, J., Caspi, A., Moffitt, T. E., Harrington, H., Milne, B. J., & Poulton, R. (2003). Prior juvenile diagnoses in adults with mental disorder: Developmental follow-back of a prospective-longitudinal cohort. *Archives of General Psychiatry, 60,* 709–717. doi:10.1001/archpsyc.60.7.709

Lupien, S. J., McEwen, B. S., Gunnar, M. R., & Heim, C. (2009). Effects of stress throughout the lifespan on the brain, behaviour and cognition. *Nature Reviews Neuroscience, 10,* 434–445. doi:10.1038/nrn2639

March, J. S., Swanson, J. M., Arnold, L. E., Hoza, B., Conners, C. K., Hinshaw, S. P., . . . Pelham, W. E. (2000). Anxiety as a predictor and outcome variable in the multimodal treatment study of children with ADHD (MTA). *Journal of Abnormal Child Psychology, 28,* 527–541. doi:10.1023/A:1005179014321

Matthys, W., Vanderschuren, L. J. M. J., Schutter, D. J. L. G., & Lochman, J. E. (2012). Impaired neurocognitive functions affect social learning processes in oppositional defiant disorder and conduct disorder: Implications for interventions. *Clinical Child and Family Psychology Review, 15,* 234–246. doi:10.1007/s10567-012-0118-7

McGowan, P. O., Sasaki, A., D'Alessio, A. C., Dymov, S., Labonté, B., Szyf, M., . . . Meaney, M. J. (2009). Epigenetic regulation of the glucocorticoid receptor in human brain associates with childhood abuse. *Nature Neuroscience, 12,* 342–348. doi:10.1038/nn.2270

McGowan, P. O., Suderman, M., Sasaki, A., Huang, T. C. T., Hallett, M., Meaney, M. J., & Szyf, M. (2011). Broad epigenetic signature of maternal care in the brain of adult rats. *PLoS ONE, 6,* e14739. doi:10.1371/journal.pone.0014739

Meaney, M. J., & Ferguson-Smith, A. C. (2010). Epigenetic regulation of the neural transcriptome: The meaning of the marks. *Nature Neuroscience, 13,* 1313–1318. doi:10.1038/nn1110-1313

Meaney, M. J., & Szyf, M. (2005). Environmental programming of stress responses through DNA methylation: Life at the interface between a dynamic environment and a fixed genome. *Dialogues in Clinical Neuroscience, 7,* 103–123.

Melas, P. A., Wei, Y., Wong, C. C. Y., Sjöholm, L. K., Åberg, E., Mill, J., . . . Lavebratt, C. (2013). Genetic and epigenetic associations of MAOA and NR3C1 with depression and childhood adversities. *International Journal of Neuropsychopharmacology, 16,* 1513–1528. doi:10.1017/S1461145713000102

Mellor, D. (2005). Normative data for the Strengths and Difficulties Questionnaire in Australia. *Australian Psychologist, 40,* 215–222. doi:10.1080/00050060500243475

Oberlander, T. F., Weinberg, J., Papsdorf, M., Grunau, R., Misri, S., & Devlin, A. M. (2008). Prenatal exposure to maternal depression, neonatal methylation of human glucocorticoid receptor gene (NR3C1) and infant cortisol stress responses. *Epigenetics, 3,* 97–106. doi:10.4161/epi.3.2.6034

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, and Computers, 32,* 396–399. doi:10.3758/BF03200807

Pluess, M., & Belsky, J. (2012). Conceptual issues in psychiatric gene-environment interaction research. *American Journal of Psychiatry, 169,* 222–223. doi:10.1176/appi.ajp.2011.11111614r

Romens, S. E., McDonald, J., Svaren, J., & Pollak, S. D. (2014). Associations between early life stress and gene methylation in children. *Child Development, 86,* 303–309. doi:10.1111/cdev.12270

Rosas, S. L. B., Koch, W., Da Costa Carvalho, M. D. G., Wu, L., Califano, J., Westra, W., . . . Sidransky, D. (2001). Promoter hypermethylation patterns of p16, O6-methylguanine-DNA-methyltransferase, and death-associated protein kinase in tumors and saliva of head and neck cancer patients. *Cancer Research, 61,* 939–942.

Schechter, J. C., Brennan, P. A., Cunningham, P. B., Foster, S. L., & Whitmore, E. (2012). Stress, cortisol, and externalizing behavior in adolescent males: An examination in the context of multisystemic therapy. *Journal of Abnormal Child Psychology, 40,* 913–922. doi:10.1007/s10802-012-9612-0

Szyf, M. (2013). The genome- and system-wide response of DNA methylation to early life adversity and its implication on mental health. *Canadian Journal of Psychiatry, 58,* 697–704.

Szyf, M., & Bick, J. (2013). DNA methylation: A mechanism for embedding early life experiences in the genome. *Child Development, 84,* 49–57. doi:10.1111/j.1467-8624.2012.01793.x

Turner, J. D., Alt, S. R., Cao, L., Vernocchi, S., Trifonova, S., Battello, N., & Muller, C. P. (2010). Transcriptional control of the glucocorticoid receptor: CpG islands, epigenetics and more. *Biochemical Pharmacology, 80,* 1860–1868. doi:10.1016/j.bcp.2010.06.037

van IJzendoorn, M. H., Bakermans-Kranenburg, M. J., & Ebstein, R. P. (2011). Methylation matters in child development: Toward developmental behavioral epigenetics. *Child Development Perspectives, 5,* 305–310. doi:10.1111/j.1750-8606.2011.00202.x

Viding, E., Hanscombe, K. B., Curtis, C. J. C., Davis, O. S. P., Meaburn, E. L., & Plomin, R. (2010). In search of genes associated with risk for psychopathic tendencies in children: A two-stage genome-wide association study of pooled DNA. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 51,* 780–788. doi:10.1111/j.1469-7610.2010.02236.x

Washburn, J. J., Romero, E. G., Welty, L. J., Abram, K. M., Teplin, L. A., McClelland, G. M., & Paskar, L. D. (2007). Development of antisocial personality disorder in detained youths: The predictive value of mental disorders. *Journal of Consulting and Clinical Psychology, 75,* 221–231. doi:10.1037/0022-006X.75.2.221

World Health Organization. (2008). *The global burden of disease: 2004 update.* Geneva, Switzerland: Author.

Child Development, September/October 2015, Volume 86, Number 5, Pages 1321–1332

Attention to Faces Expressing Negative Emotion at 7 Months Predicts Attachment Security at 14 Months

Mikko J. Peltola and Linda Forssman
University of Tampere

Kaija Puura
University of Tampere and Tampere University Hospital

Marinus H. van IJzendoorn
Leiden University

Jukka M. Leppänen
University of Tampere

To investigate potential infant-related antecedents characterizing later attachment security, this study tested whether attention to facial expressions, assessed with an eye-tracking paradigm at 7 months of age ($N = 73$), predicted infant–mother attachment in the Strange Situation Procedure at 14 months. Attention to fearful faces at 7 months predicted attachment security, with a smaller attentional bias to fearful expressions associated with insecure attachment. Attachment disorganization in particular was linked to an absence of the age-typical attentional bias to fear. These data provide the first evidence linking infants' attentional bias to negative facial expressions with attachment formation and suggest reduced sensitivity to facial expressions of negative emotion as a testable trait that could link attachment disorganization with later behavioral outcomes.

Developing during the 1st year, the attachment behavioral system facilitates different response patterns upon threatening situations and separation from caregivers (Bowlby, 1969). Compared to securely attached infants who readily seek and manage to get comfort from the caregiver when feeling frightened, insecurely attached infants are characterized by an inhibition of outward signs of distress and apparent indifference toward the caregiver (avoidance), clinginess and anger toward the caregiver (resistance), or a lack of a coherent behavioral strategy and potential signs of fear of the caregiver (disorganization; Ainsworth, Blehar, Waters, & Wall, 1978; Main & Solomon, 1990). An extensive amount of research has established the importance of sensitive and supportive caregiving on the emergence of secure attachment (Ainsworth et al., 1978; Bernier, Matte-Gagné, Bélanger, & Whipple, 2014; De Wolff & van IJzendoorn, 1997), and provided evidence for an enduring impact of attachment security on a range of measures of social adaptation such as externalizing and internalizing problems, social competence, and emotion regulation (Fearon, Bakermans-Kranenburg, van IJzendoorn, Lapsley, & Roisman, 2010; Groh, Roisman, van IJzendoorn, Bakermans-Kranenburg, & Fearon, 2012; Groh et al., 2014; Sroufe, 2005).

Although the 1st year of life is critical for attachment formation and the role of the caregiving environment in this process has been investigated in detail, there are considerably fewer prospective studies investigating how developmental trajectories during this period differ in infants later classified as securely or insecurely attached. Studies addressing this question have shown that child characteristics such as temperament (Van IJzendoorn & Bakermans-Kranenburg, 2012) and candidate genetic variants (e.g., Luijk et al., 2011) may not consistently predict attachment, but there is some evidence for responses more closely related to interactive contexts, such as infants' positive affect

---

We gratefully acknowledge the efforts of the families who participated in this project, and Henna Salovaara, Mari Fadjukoff, Mia Huolman, Susanna Keinänen, Iiris Ruotsalainen, Milla Miettinen, Henriikka Nisula, and Maria Kataja for their help in data collection. This research was supported by a grant from the Kone Foundation and a postdoctoral fellowship from the Institute for Advanced Social Research, University of Tampere, for Mikko J. Peltola; an award from the the Netherlands Organization for Scientific Research (SPINOZA prize) for Marinus H. van IJzendoorn; and a grant from the European Research Council (Grant 283763) for Jukka M. Leppänen.

Correspondence concerning this article should be addressed to Mikko J. Peltola, Human Information Processing Laboratory, School of Social Sciences and Humanities, University of Tampere, 33014 Tampere, Finland. Electronic mail may be sent to mikko.peltola@uta.fi.

© 2015 The Authors.
Child Development published by Wiley Periodicals, Inc. on behalf of Society for Research in Child Development.
This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.
All rights reserved. 0009-3920/2015/8605-0002
DOI: 10.1111/cdev.12380

expressions and behavioral responses during the Still Face procedure, to predict later attachment (e.g., Braungart-Rieker et al., 2014; Mesman, van IJzendoorn, & Bakermans-Kranenburg, 2009, for a review). In the present study, we continued this line of research by testing a hypothesis that infants' attention to facial emotion expressions, particularly to emotions of negative valence, is associated with later attachment security.

There are good reasons to assume that attachment formation in infants is marked by changes in attention to facial expressions. Models of attachment and socioemotional information processing suggest that recurrent attachment-related experiences shape representational, physiological, and behavioral responses to emotional information, and to cues signaling threat in particular (Bowlby, 1969; Cassidy, Jones, & Shaver, 2013; Dykas & Cassidy, 2011; Vrtička & Vuilleumier, 2012). Cues signaling threat can range from direct and attachment-specific threat signals (such as the separation from or displays of anger by the caregiver) to more ambiguous cues of potential danger (such as fearful faces). It is argued that attachment-related information processing biases toward potential threat cues in the environment are constantly operating (cf. Bowlby, 1969; Dykas & Cassidy, 2011), although experiencing threat oneself may amplify the preexisting response patterns. Although there is evidence from adults showing that attachment security is associated with patterns of attending toward or away from threat-related stimuli such as angry faces (e.g., Dewitte & De Houwer, 2008), no studies have investigated the early origins of attachment-related information processing biases in infancy. The possibility that variations in attention to facial expressions of negative emotion are associated with the evolving attachment relationship in early development is further suggested by findings showing that individual variations in attention to fearful facial expressions in 7-month-old infants are related to maternal stress and depressive symptoms (Forssman et al., 2014) and that attention to angry faces is enhanced in school-age children with histories of parental maltreatment (e.g., Pollak & Kistler, 2002).

Research documenting the early ontogeny of the perception of emotional stimuli has shown that during the second half of the 1st year, infants begin to reliably discriminate between a range of emotional expressions presented in different channels such as the face (Leppänen & Nelson, 2009; Nelson & Dolgin, 1985), voice (Flom & Bahrick, 2007), and body movements (Zieber, Kangas, Hock, & Bhatt, 2014). A notable developmental feature of emotion perception at around 5–7 months of age is a robust attentional bias toward facial expressions of fear. This bias is typically observed in prolonged engagement of attention to fearful facial expressions in comparison to happy or neutral stimuli (Nakagawa & Sukigara, 2012; Nelson & Dolgin, 1985; Peltola, Hietanen, Forssman, & Leppänen, 2013; Peltola, Leppänen, Palokangas, & Hietanen, 2008), augmented heart rate deceleration (Leppänen et al., 2010; Peltola, Leppänen, & Hietanen, 2011), and attention-related electrocortical brain responses to fearful faces (Leppänen, Moulson, Vogel-Farley, & Nelson, 2007; Nelson & de Haan, 1996; Peltola, Leppänen, Mäki, & Hietanen, 2009). Importantly, the attention bias appears to be relatively specific to fearful expressions as similar biases have not been observed to sad or angry expressions (e.g., Grossmann, Striano, & Friederici, 2007; Soken & Pick, 1999) or to faces displaying novel, nonfearful expressions (Peltola, Leppänen, Vogel-Farley, Hietanen, & Nelson, 2009; Peltola et al., 2008), although the extent to which infants assign specific emotional meaning (e.g., fear) to pictures of facial expressions remains debatable (Kagan & Herschkowitz, 2005). The robustness and relative specificity of infants' attentional bias to fearful facial expressions are reasons to assume that this phenomenon reflects an important aspect of early emotional information processing, and variability in this bias may therefore present a particularly suitable starting point for investigating the early antecedents of later attachment. However, as yet, there have been no studies examining the potential of interindividual variation in attentional biases to negative facial expressions as a marker of subsequent social and emotional development, or as a precursor of more refined functional behavioral responding to emotional events (cf. Walle & Campos, 2012).

In the present longitudinal study we examined whether infants' attentional biases to facial expressions at 7 months of age predict mother–infant attachment quality after 1 year of age. Infants were tested with an eye-tracking measure of attention to facial expressions (i.e., the Overlap paradigm) at 7 months of age and mother–infant attachment was assessed at 14 months of age with the Strange Situation Procedure (SSP; Ainsworth et al., 1978). The Overlap paradigm (modified from the original paradigm by Aslin & Salapatek, 1975), examines infants' gaze shifts from a central stimulus (i.e., a neutral, happy, or fearful expression, or a face-shaped control stimulus) to a high-contrast stimulus in the left or right periphery. Due to the temporally overlapping presentation of the two stimuli, a gaze shift to the peripheral stimulus requires active disengagement of attention from the central face stimulus (Colombo, Brez, & Curtindale, 2012), which makes this task suitable for probing differences in attention allocation between distinct centrally presented stimuli. Attentional biases can be assessed most reliably by calculating the probability of attention shifts from the centrally presented stimuli toward the peripheral stimuli, and a robust bias to fearful expressions has been observed as a higher probability of missing attention shifts from fearful versus happy/neutral faces in independent infant samples (Forssman et al., 2014; Leppänen et al., 2011; Nakagawa & Sukigara, 2012; Peltola et al., 2008, 2011).

Our primary aim was to test for differences in attention to facial expressions between infants grouped as securely or insecurely attached (i.e., avoidant, resistant, and disorganized), with additional exploratory analyses aiming at a further characterization of the data across the attachment subgroups and with a continuous measure of attachment disorganization. As described above, we were particularly interested in whether early attentional responses to negative facial expressions predict the activation of attachment behaviors in later development. Because previous studies have shown infants within this age range to display robust attentional biases to fearful faces, we selected fearful faces as the negative emotion category in this study. There are no prior studies examining the association between processing of facial expressions and attachment security in infants, which led us to adopt an exploratory approach and refrain from strong directional hypotheses. Theoretical and empirical work with adults (Dewitte & De Houwer, 2008; Dykas & Cassidy, 2011) suggest that, possibly serving a regulatory function to suppress emotional overarousal, a relatively automatic tendency to divert attention away from threat-related cues is associated with insecure attachment. In the present paradigm such tendency would be observed as a relatively smaller attention bias to fearful expressions. Alternatively, the reverse could also be expected; that is, secure attachment could be associated with a relatively smaller attention bias to fearful expressions, possibly due to a decreased need to monitor the environment for potential signs of threat in infants with a more established expectation of caregiver availability.

Besides infants' attention to facial expressions, maternal sensitivity was assessed when the infants were 7 months old. Maternal sensitivity was assessed to examine whether attentional biases to facial expressions, and their potential associations with later attachment, are associated with differences in the mothers' ability to detect and respond to infant signals. Apart from studies with maternal personality (de Haan, Belsky, Reid, Volein, & Johnson, 2004) or depressive symptoms (Forssman et al., 2014) as a proxy of the caregiving environment, there is a lack of research on the associations between parental caregiving sensitivity and the processing of emotional stimuli in infancy. Also, as parental sensitivity has an established role in the development of attachment security, it is important to determine whether other potential predictive factors (such as attentional biases to facial expressions) are independent from or interact with parental sensitivity in predicting later attachment security. We expected secure attachment to be associated with higher maternal sensitivity, but did not advance a specific directional hypothesis with respect to the influence of maternal sensitivity on attention to facial expressions.

Method

Participants

The infants were recruited from an ongoing longitudinal cohort study ($N = 126$; 44% females; Forssman et al., 2014; Peltola et al., 2013) in which eye-tracking assessments of attention to facial expressions with the Overlap paradigm have been conducted at 5 months ($M = 152.43$ days, $SD = 3.64$) and 7 months of age ($M = 213.85$ days, $SD = 4.39$). All infants were healthy, full-term (≥ 37 weeks), and from urban, middle-class families of Caucasian ethnicity. A total of 73 infants from the original cohort participated in the attachment assessment at 14 months of age ($M = 414.85$ days, $SD = 20.24$; 44% females). The notable loss of infants for the follow-up attachment assessments is partly explained by the fact that the attachment assessment was introduced as a supplemental visit after the onset of the longitudinal study. Importantly, no statistical differences were found between those who did and did not participate in the attachment assessment in maternal sensitivity, attention to facial expressions, or temperament (as measured with the Infant Behavior Questionnaire; Rothbart, 1981), *ps* = .28–.96, indicating that the participants included in the current analyses are representative of the complete longitudinal study cohort. All infants with ≥ 3 scorable trials in each of the four stimulus conditions in the Overlap paradigm and attachment data were included in the present analyses. Given that a relatively greater number of infants had insufficient eye-tracking data in the 5-month assessment ($n = 26$) as compared to the 7-month assessments ($n = 11$), the final analyses reported here are based on infants who had valid eye-tracking data from the 7-month-assessment and had participated in the attachment assessments at 14 months of age ($n = 62$; 39% females; see Data S1 in the Online Supporting Information for the 5-month data analyses). Approval for the project was obtained from the Ethical Committee of Tampere University Hospital, and informed written consent was obtained from the parent of each child.

Measures

Attention to Facial Expressions

Eye-tracking data were collected during the Overlap paradigm for measuring attention to facial expressions and face-shaped control stimuli (Figure 1). The facial expression stimuli were neutral, happy, and fearful expressions posed by two female models, and the face-shaped control stimuli were phase-scrambled images of both of the models' faces, preserving the outer contour of the face. The face and control stimuli measured 15° and 11° vertically and horizontally, respectively. The session started with the placement of an electrode net for recording electroencephalography (for these data, see Yrttiaho, Forssman, Kaatiala, & Leppänen, 2014). During the experiment, the infants sat on their parent's lap at a 60-cm viewing distance from a corneal-reflection eye-tracker monitor with a 58-cm diameter (Tobii TX300; Tobii Technology, Stockholm, Sweden), which recorded data on infants' eye positions on the screen with a 300-Hz sampling rate. Data collection started with the calibration of the eye-tracker cameras during which an audiovisual animation was presented sequentially at every corner and the center of the screen. If the first calibration was not successful (i.e., one or more calibration points were missing or were not properly calibrated), the calibration was repeated at least two times to attain satisfactory calibration for all five locations. If one or more calibration points were missing after more than 2 recalibration attempts, the final calibration outcome was accepted and the experiment was started. If none of the locations were successfully calibrated after several recalibration attempts, the experiment was then run without collecting eye-tracking data. In the Overlap paradigm, a fixation stimulus (a red circle repeatedly expanding from 0.4° to 4.3°) preceded each trial. Once the infant looked at the fixation stimulus, the experimenter pressed a key to present a face or a face-shaped control stimulus on the center of the screen. The central stimulus was flanked after 1,000 ms by a peripheral stimulus presented 14° equiprobably on the left or right for 3,000 ms. The peripheral stimuli were black-and-white vertically arranged circles or a checkerboard pattern, measuring 15° × 4° visual angle. Each stimulus condition (i.e., control, neutral, happy, fearful) was repeated six times for both models in a random order (i.e., a total of 48 trials), with the constraint that the same stimulus condition was presented no more than four times in a row and the peripheral stimulus on the same side no more than three times in a row. Stimulus presentation was controlled by E-Prime 2 software (Psychology Software Tools Inc., www.pstnet.com).

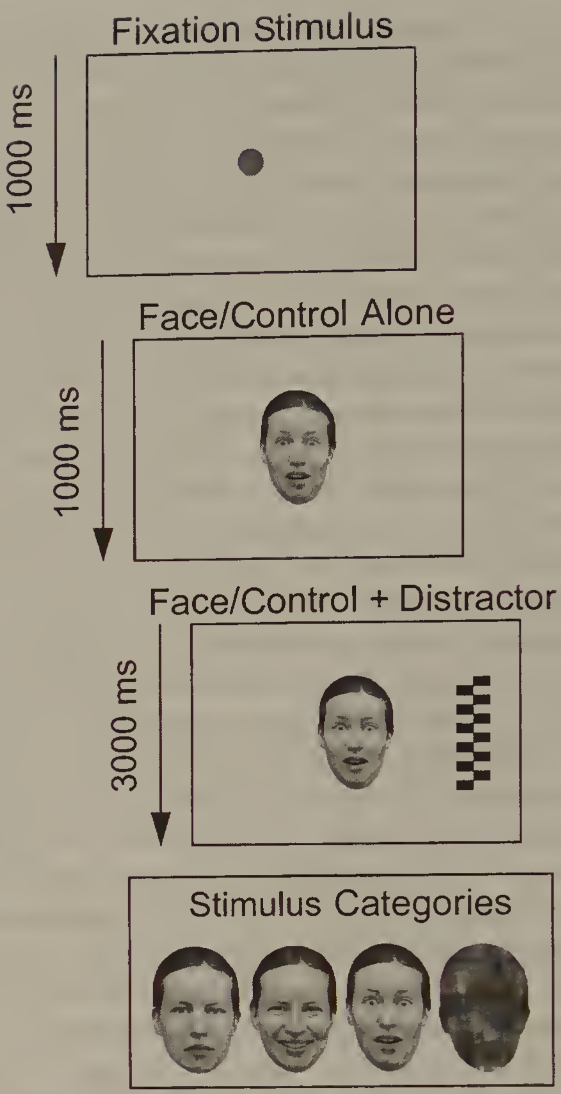


Figure 1. The sequence of events in the Overlap paradigm and examples of the stimuli.

All data processing from the raw x–y gaze position coordinates to the parameters reflecting attention disengagement were completed automatically by using *gazeAnalysisLib*, a library of MATLAB

(Mathworks, Natick, MA) routines for offline analysis of raw gaze data (Leppänen, Forssman, Kaatiala, Yrttiaho, & Wass, 2014). First, a 37-sample median filter was applied for removing abrupt spikes in the gaze data (attributable to technical artifacts). Second, data segments with a maximum of 200 ms of missing eye position data were interpolated by continuing the last recorded *x*- and *y*-coordinates until the tracking came back online. Third, bad trials were excluded from further analyses. These included trials with infant movement, inattentiveness, or technical problems in eye tracking (fragmented tracking), resulting in more than 200 ms of missing eye position data during the trial or the infant looking at the central stimulus for less than 70% of the time preceding gaze disengagement or the end of the analysis period (on average 21.9% of all trials). Additionally, a small number of trials were excluded due to anticipatory eye movements (i.e., a saccade commenced less than 150 ms after the peripheral stimulus onset; 0.5% of all trials) or computer timing errors (i.e., the stimuli not presented for the correct duration; 0.15% of all trials). No significant difference was observed in the number of excluded trials between securely and insecurely attached infants, $t(68) = 1.11$, $p = .27$. Of the remaining scorable trials, the proportion of *missing attention shifts* (i.e., no eye movement toward the peripheral stimulus during a 150- to 1,000-ms time window after the peripheral stimulus onset) was calculated for each stimulus condition (cf. Leppänen et al., 2011). In addition, *attentional bias scores* were calculated for each stimulus type for post hoc comparisons. The bias scores are deviation contrasts, which reflect the relative weighting of attention to particular face stimuli in the context of other stimuli (e.g., missing attention shifts for fear minus the mean of missing attention shifts for all other face stimuli). To be included in the statistical analyses a minimum of three scorable trials was required for each stimulus condition. On average, the infants included in the analyses had 9.2, 9.4, 9.0, and 9.4 scorable trials in the control, neutral, happy, and fearful stimulus conditions, respectively, with no differences in the number of scorable trials between securely and insecurely attached infants, $F(1, 60) = .48$, $p = .49$.

Maternal Sensitivity

Mother–infant dyads were videotaped at their homes during a free-play interaction for 15 min within 2 weeks after the 7-month laboratory session. Maternal sensitivity was assessed from videotapes using the Emotional Availability Scales (EAS), 4th ed. (Biringen, 2008) by a certified coder (author 3), unaware of infants' attachment classifications and performance in the eye-tracking task. Maternal sensitivity was coded on a scale of 1–7, consisting of parental affect expression, the ability to detect and appropriately react to child signals and behaviors, and communicating in respectful ways. The EAS data were available for 69 (of 73) mother–infant dyads. The average sensitivity score within this sample was 5.14 ($SD = 1.14$, skewness = .20), which is similar to nationally representative maternal sensitivity data (e.g., Merras-Salmio et al., 2013). Intercoder agreement (intraclass correlation [ICC]) between the main coder and an independent certified coder on a set of 20 cases was ICC = .95.

Infant–Mother Attachment

Infants and mothers were observed in the SSP (Ainsworth et al., 1978). In accordance with standard procedures, the SSP consisted of seven 3-min episodes, including two separations from and two reunions with the mother, and interaction with a female stranger. Infants' attachment behaviors were coded from videotapes according to the organized attachment scales by Ainsworth et al. (1978) and the Main and Solomon (1990) coding system for assessing attachment disorganization. On the basis of ratings on 7-point scales assessing infants' proximity seeking, contact maintenance, resistance, and avoidance during the two reunion episodes, infants were first classified as secure (B), insecure-avoidant (A), or insecure-resistant (C). Next, signs of attachment disorganization (D) during each episode when the mother was present were rated using a scale from 1 (*no signs of disorganization*) to 9 (*strong signs of disorganization*), with scores higher than 5 receiving a disorganized classification. Infant behaviors contributing to attachment disorganization scoring include contradictory behaviors (e.g., rapid avoidance following a cry for the mother), stereotypical or anomalous behaviors, stilling or freezing, direct signs of apprehension regarding the mother, and misdirected or disoriented behaviors. Intercoder agreement was calculated from 18% of the sample by two coders (authors 1 and 4, both unaware of the maternal sensitivity scores or infants' performance in the eye-tracking task). For the four-way ABCD classifications, presence versus absence of secure attachment (i.e., B vs. non-B), and the presence versus absence of attachment disorganization (i.e., D vs. non-D), intercoder agreement was 92%

(κs = .88, .85, and .81, respectively). Agreement on the continuous D scores was ICC = .91. For the remaining sample, the ABCD classifications were based on a consensus between the first author and expert coders from Leiden University. In the full sample of 73 infants, 46 infants (63%) were classified as securely attached (B), 10 (14%) as insecure-avoidant (A), 6 (8%) as insecure-resistant (C), and 11 (15%) as disorganized (D). Within the sample of 62 infants having sufficient eye-tracking data at 7 months, the distribution was: B = 37 (60%), A = 8 (13%), C = 6 (10%), and D = 11 (17%). These distributions were thus highly concordant with meta-analytic data on attachment distributions in Western samples (e.g., Van IJzendoorn, Schuengel, & Bakermans-Kranenburg, 1999), $\chi^2 < .58$, $p = .90$. The statistical analyses were conducted using the secure versus insecure grouping and the continuous D scores as dependent variables.

Results

First, an analysis of variance of the missing attention shift data at 7 months was conducted to test whether the infants included in the present sample show a similar pattern of attention to faces as has been observed in previous research. Replicating previous findings, the main effect of face stimulus was significant, $F(3, 183) = 41.34$, $p < .0001$, partial $\eta^2 = .40$, due to the proportion of missing attention shifts being significantly higher to fearful faces ($M = .52$; $SD = .25$) than to happy ($M = .36$, $SD = .25$) and neutral ($M = .36$, $SD = .25$) faces, or to the control stimuli ($M = .18$, $SD = .21$), all Bonferroni-corrected comparisons $p < .0001$, $d > .64$. Other preliminary analyses showed that missing attention shifts to faces at 7 months of age were not associated with gender or maternal sensitivity (ps = .16–.99). Attachment grouping (secure vs. insecure) was not associated with maternal sensitivity, $t(67) = .45$, $p = .66$, or age (in days) at attachment assessment, $t(71) = .49$, $p = .63$. There were more males than females with an insecure classification (20 vs. 7) as compared to the gender distribution within the secure group (21 vs. 25), $\chi^2 = 5.58$, $p = .018$. The continuous D scores were not associated with maternal sensitivity, $r = .05$, $p = .71$; age at attachment assessment, $r = .03$, $p = .79$; or gender, $t(71) = .40$, $p = .69$.

In the main analyses, logistic regression analysis was conducted to predict secure versus insecure attachment grouping and linear regression was used to predict attachment disorganization (the continuous D scores) on the basis of missing attention shift scores to faces and face-shaped control stimuli at 7 months of age (with the predictor variables mean centered). The regression analyses were first run with gender and maternal sensitivity included as predictor variables, but as no interactions involving these variables emerged (i.e., gender and maternal sensitivity did not interact with missing attention shift scores to predict attachment security or disorganization), they were excluded from the final analyses to reduce the number of predictors and to improve model fit.

The logistic regression model with attachment grouping (secure vs. insecure) as the dependent variable and missing attention shift scores as predictors was significant, $\chi^2 = 10.40$, $p = .034$, $R^2 = .15$. Attachment grouping was predicted significantly by the proportion of missing attention

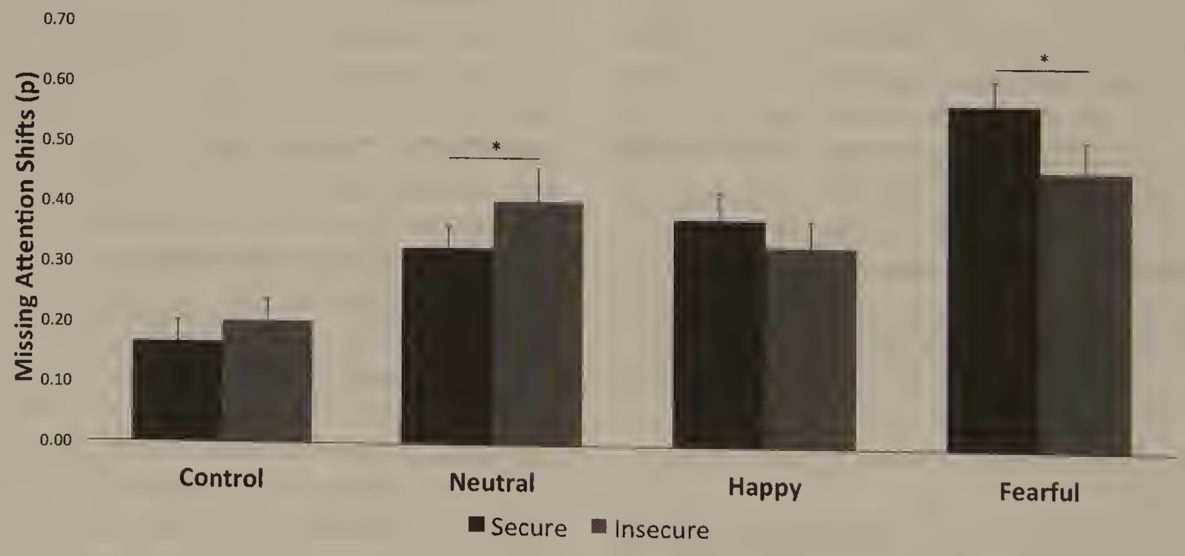


Figure 2. Mean proportion (p) of missing attention shifts in different stimulus conditions in 7-month-old infants, grouped by attachment security. Error bars represent the standard error of mean.
$^*p < .05$.

shifts to neutral ($B = -3.90$, $SE = 1.79$, $p = .03$) and fearful ($B = 3.39$, $SE = 1.61$, $p = .035$) faces (Figure 2). Separate *t* tests with the attentional bias scores showed that secure attachment was associated with a larger attentional bias to fearful faces, $t(60) = 2.59$, $p = .012$, $d = .67$, while the bias scores to neutral faces were larger in insecure than secure infants, $t(60) = 2.46$, $p = .017$, $d = .64$. As an exploratory analysis to estimate how these effects manifest in the different insecure subgroups, a multinomial logistic regression with the four-way attachment classification (A, B, C, D) as the dependent variable and missing attention shift scores as predictors was run. The model was significant, $\chi^2 = 23.40$, $p = .024$, $R^2 = .31$. Inspection of the parameter estimates (with secure attachment as the reference category) indicated that attachment avoidance was associated with a relatively larger bias to neutral ($p = .023$) but not fearful ($p = .79$) faces, whereas both resistant and disorganized attachment were associated with a smaller bias to fearful faces (*p*s = .039 and .012, respectively), but did not differ from secure attachment in attention to neutral faces (*p*s = .44 and .15, respectively).

The association with attachment disorganization was further tested with a linear regression analysis with the D score as the dependent variable and missing attention shift scores at 7 months as predictors. The model was significant, $F(4, 61) = 2.68$, $p = .041$, $R^2 = .16$, with attention to fearful faces as the only significant predictor of the D scores, $\beta = -.48$, $p = .004$. Finally, to test whether the attentional bias to fearful faces is equally strongly associated with disorganization and resistance, a partial correlation between fear bias scores and D scores was calculated while controlling for resistance scores (averaged across the two reunions during the SSP, with higher scores on a 7-point scale indicating higher resistance), and vice versa (i.e., correlating fear bias and resistance scores while controlling for D scores). This analysis showed that the negative association between fear bias and attachment disorganization ($r = -.35$, $p = .006$, i.e., higher D scores associated with a smaller attentional bias to fearful faces; see Figure 3) remained significant even while controlling for resistance scores ($r = -.26$, $p = .046$), whereas the correlation between fear bias and resistance ($r = -.28$, $p = .03$) became nonsignificant when the D scores were taken into account ($r = -.14$, $p = .28$), indicating that the decreased attentional bias to fearful faces was more strongly associated with attachment disorganization than resistance. A one-sample *t* test with the subgroup classified as disorganized further indicated an absence of fear bias in these infants, as the magnitude of their fear

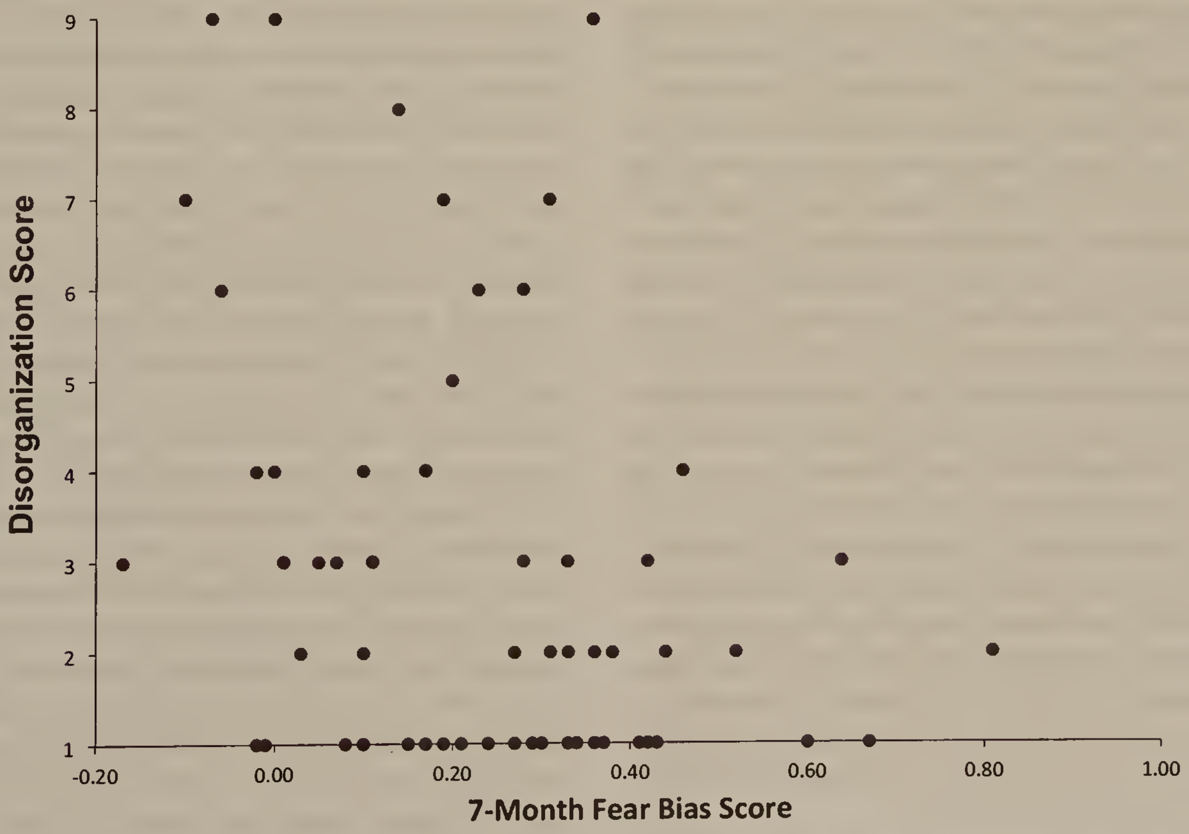


Figure 3. The correlation ($r = -.35$, $p = .006$) between attachment disorganization and attentional bias to fearful faces at 7 months of age.

bias ($M = .09$) did not differ from zero, $t(10) = 1.59$, $p = .14$.

Discussion

In the present study, attention to facial expressions at 7 months of age predicted attachment security at 14 months of age: As compared to securely attached infants, insecurely attached infants showed a smaller attentional bias to fearful faces and a larger bias to neutral faces (although the latter effect appeared to be restricted to the small subgroup of avoidantly attached infants). Further exploratory analyses showed that the smaller attentional bias to fearful faces was most clearly associated with increasing signs of attachment disorganization.

Together, these findings are consistent with the hypothesis that variations in attention to negative emotions, and possibly to cues signaling threat in particular, are associated with attachment formation and may provide a useful marker of infants' attachment status prior to the standard observational assessment of attachment after the first birthday. Regarding the potential mechanisms mediating this association, the results could be seen as consistent with the model of Dykas and Cassidy (2011) proposing that an acquired tendency to divert attention away from threat-related cues and suppress emotional overarousal is a marker of insecure attachment. It is unclear, however, whether the seemingly similar pattern of results implies that the same mechanisms (i.e., suppression of arousal) are functional in infants and adults, or whether other factors such as lower or even blunted sensitivity to the emotional signal conveyed by fearful faces (and potentially other expressions of negative valence) in insecurely attached infants are more likely responsible for the observed effects. Interestingly, our supplementary analyses showed that variations in attention bias at 5 months were not associated with later attachment status and that an increase in the attentional bias to fearful expressions from 5 to 7 months was more pronounced in securely attached infants than insecure and disorganized infants (see Data S1). Although tentative, these results raise the possibility that individual variations in the developmental emergence of an attentional bias to faces expressing negative emotions (such as fearful expressions) between 5 and 7 months of age are particularly informative for understanding attachment formation (cf. Leppänen & Nelson, 2012).

Whereas a robust attentional bias to fearful faces was related to secure attachment and low attachment disorganization scores, higher levels of attachment disorganization were characterized by a diminished attentional bias to fearful faces in this study. The absence of the age-typical (e.g., Forssman et al., 2014; Peltola et al., 2008) attentional bias to fearful expressions in disorganized infants is highly interesting in light of data on the associations between attachment disorganization in infancy and later emotional and behavioral outcomes. Meta-analytic data indicate that while attachment disorganization shows only a negligible association with later internalizing symptoms (e.g., anxiety and depression; Groh et al., 2012), disorganized attachment is associated with later externalizing symptoms (e.g., aggressive and oppositional behavior; see Fearon et al., 2010). Facets of externalizing symptomatology such as callous-unemotional traits and antisocial behavior have been linked with specific impairments in sensitivity to other individuals' fearfulness rather than a generalized impairment in processing emotion expressions (Dadds et al., 2006; Marsh & Blair, 2008), and conversely, higher sensitivity to detect fearful cues is related to higher prosocial tendencies (Marsh, Kozak, & Ambady, 2007). It is intriguing to hypothesize that a lower sensitivity to the distress signals of others could provide a remarkably early-emerging endophenotypic marker relating attachment disorganization in infancy to later externalizing problems.

What is not fully revealed by the current results is the causal direction of associations, that is, whether variation in attentional biases to emotional stimuli is an inherent characteristic of the infant imposing independent effects on the emergence of attachment security, or whether attentional biases are shaped by the caregiving environment at 7 months of age and could be more readily interpreted as an early correlate of the attachment relationship rather than an independent determinant. Partly related to this question, Tharner et al. (2011) showed that attachment disorganization at 14 months was predicted by anatomical variation in 6-week-old infants' subcortical structures (i.e., the diameter of the gangliothalamic ovoid). As this structure is linked to the limbic areas subserving the processing of emotional information, such findings indicate potentially inherent neurobiological factors that may predispose infants to be differentially sensitive to emotional stimuli. Although such findings and the absence of associations between attentional biases and maternal sensitivity in the present study could be seen to favor the independent role of attentiveness to emotional signals on determining later attachment, we hesitate to draw such strong conclusions given potential limitations in our assays of maternal sensitivity. For example, the association between attachment security and sensitivity was not found in this study although it has been extensively documented meta-analytically (De Wolff & van IJzendoorn, 1997) and with behavior genetic (Fearon et al., 2006) and intervention (Bakermans-Kranenburg, van IJzendoorn, & Juffer, 2003) designs. Multiple factors have likely contributed to this null effect, including modest sample size, limited variance in maternal sensitivity scores, and the relatively unconstrained free-play observation, as it has been proposed that sensitivity to infant distress signals in stress-inducing contexts provides a better estimate of the influence of sensitivity on attachment than nondistressing assessments (Leerkes, 2011). Furthermore, apart from the global assessment of caregiving sensitivity used in the present study, the influence of more focal interactive behaviors on infants' processing of emotional signals has not been investigated. Specific facets of interactive behavior such as atypical parental mirroring of infant affect (DeOliveira, Bailey, Moran, & Pederson, 2004; Gergely, 2004) as well as frightened and frightening emotion displays toward the infant (Main & Hesse, 1990; Schuengel, Bakermans-Kranenburg, & van IJzendoorn, 1999) have been related to the development of attachment disorganization, and suggested to be partly mediated by infants' ability to recognize their own and others' emotions (DeOliveira et al., 2004). A potentially fruitful agenda for future studies is to investigate in more detail how such focal aspects of parenting are associated with infants' perceptual processing of emotion signals.

In a wider context, the present results showing that infants' attentional bias to fearful faces is associated with an apparently positive developmental outcome (secure attachment) are of interest as previous studies have linked attentional biases to threat-related stimuli (such as angry and fearful faces) with exposure to stress (Forssman et al., 2014) and proneness to anxiety (Bar-Haim, Lamy, Pergamin, Bakermans-Kranenburg, & van IJzendoorn, 2007). First, the present results suggest that the absence of age-typical attention biases, not only heightened levels of this bias, may be informative as a marker of early childhood development. Second, the present results call for caution in interpreting infants' attentional biases to threat-related stimuli as "risk factors," even though this interpretation appears plausible in light of several studies in adults linking the bias with anxiety (Bar-Haim et al., 2007). In particular, the present results raise the possibility that the presence of threat-related attentional biases in infancy is an important and potentially transient (see Peltola et al., 2013) aspect of typical socioemotional development. The results also call for further work to examine infants' attentional biases to fearful faces and other threat-related stimuli not only in the context of ordinary variations in parental depression and stress symptoms (Forssman et al., 2014) but also under more severe conditions.

Limitations

In the future, the links between attachment and social information processing need to be investigated in ethnically and socioeconomically more diverse samples to assess the generalizability of the effects (the present study being restricted to a Caucasian, middle-class sample). Likewise, although the association between attentional biases and later attachment security was not influenced by gender in this study, a replication of these findings in a larger sample without boys being overrepresented in the insecure category is needed. A larger sample will also be important to enable more reliable analyses of the patterns of attentional bias within each insecure attachment subgroup. Finally, it is important to stress that the present results do not unequivocally permit interpreting the observed attention effects to be specific to fearful expressions instead of a more general bias to threat-related or negatively valenced emotion expressions, and a further investigation of the scope of the observed effects will be important. Considering the idea that attachment representations impose a generalized influence on the deployment of processing resources toward potential threats in the environment (e.g., Dewitte & De Houwer, 2008; Dykas & Cassidy, 2011), it could be expected that the observed effects replicate with other threat-related stimuli such as expressions of anger directed toward the infant. We believe, however, that fearful expressions are a particularly suitable class of stimuli to start investigating the links between attachment and processing of emotional stimuli in infancy as there is currently no evidence for a similarly robust attentional bias to other negatively valenced facial expressions, such as anger or sadness in infants in the same age range (e.g., Grossmann et al., 2007; Soken & Pick, 1999).

Conclusions

Notwithstanding the limitations of the present study, the investigation of infants' attention to facial expressions as a precursor of attachment represents a novel approach to the study of early attachment relationships and adheres to calls to integrate experimental paradigms and measurement tools from other domains of developmental research to the study of socioemotional development in infancy (Olson & Dweck, 2009). The main finding was a smaller attentional bias to fearful faces in insecure versus secure infants, and this effect was most clearly associated with attachment disorganization. While we consider this as an initial finding in need of a replication in a larger sample and with a broader range of negative emotional expressions, it is consistent with proposed models of attachment and social information processing (Dykas & Cassidy, 2011) and suggests altered sensitivity to threat-related cues in infancy as a testable trait linking attachment disorganization to later behavioral outcomes.

References

Ainsworth, M. D. A., Blehar, M. C., Waters, E., & Wall, S. (1978). *Patterns of attachment: A psychological study of the strange situation*. Hillsdale, NJ: Erlbaum.

Aslin, R. N., & Salapatek, P. (1975). Saccadic localization of visual targets by the very young human infant. *Perception & Psychophysics, 17*, 293–302. doi:10.3758/BF03203214

Bakermans-Kranenburg, M. J., van IJzendoorn, M. H., & Juffer, F. (2003). Less is more: Meta-analyses of sensitivity and attachment interventions in early childhood. *Psychological Bulletin, 129*, 195–215. doi:10.1037/0033-2909.129.2.195

Bar-Haim, Y., Lamy, D., Pergamin, L., Bakermans-Kranenburg, M. J., & van IJzendoorn, M. H. (2007). Threat-related attentional bias in anxious and nonanxious individuals: A meta-analytic study. *Psychological Bulletin, 133*, 1–24. doi:10.1037/0033-2909.133.1.1

Bernier, A., Matte-Gagné, C., Bélanger, M., & Whipple, N. (2014). Taking stock of two decades of attachment transmission gap: Broadening the assessment of maternal behavior. *Child Development, 85*, 1852–1865. doi:10.1111/cdev.12236

Biringen, Z. (2008). *Emotional Availability (EA) scales* (4th ed.). Retrieved from http://www.emotionalavailability.com

Bowlby, J. (1969). *Attachment and loss: Attachment* (Vol. 1). New York, NY: Basic Books.

Braungart-Rieker, J. M., Zentall, S., Lickenbrock, D. M., Ekas, N. V., Oshio, T., & Planalp, E. (2014). Attachment in the making: Mother and father sensitivity and infants' responses during the still-face paradigm. *Journal of Experimental Child Psychology, 125*, 63–84. doi:10.1016/j.jecp.2014.02.007

Cassidy, J., Jones, J. D., & Shaver, P. R. (2013). Contributions of attachment theory and research: A framework for future research, translation, and policy. *Development and Psychopathology, 25*, 1415–1434. doi:10.1017/S0954579413000692

Colombo, J., Brez, C. C., & Curtindale, L. M. (2012). Infant perception and cognition. In R. Lerner, A. Easterbrooks, J. Mistry, & I. Weiner (Eds.), *Handbook of psychology: Vol. 6. Developmental psychology* (2nd ed., pp. 61–89). Hoboken, NJ: Wiley.

Dadds, M. R., Perry, Y., Hawes, D. J., Merz, S., Riddell, A. C., Haines, D. J., . . . Abeygunawardane, A. I. (2006). Attention to the eyes and fear-recognition deficits in child psychopathy. *British Journal of Psychiatry, 189*, 280–281. doi:10.1192/bjp.bp.105.018150

de Haan, M., Belsky, J., Reid, V., Volein, A., & Johnson, M. H. (2004). Maternal personality and infants' neural and visual responsivity to facial expressions of emotion. *Journal of Child Psychology and Psychiatry, 45*, 1209–1218. doi:10.1111/j.1469-7610.2004.00320.x

DeOliveira, C. A., Bailey, H. N., Moran, G., & Pederson, D. R. (2004). Emotion socialization as a framework for understanding the development of disorganized attachment. *Social Development, 13*, 437–467. doi:10.1111/j.1467-9507.2004.00276.x

Dewitte, M., & De Houwer, J. (2008). Adult attachment and attention to positive and negative emotional face expressions. *Journal of Research in Personality, 42*, 498–505. doi:10.1016/j.jrp.2007.07.010

De Wolff, M. S., & van IJzendoorn, M. H. (1997). Sensitivity and attachment: A meta-analysis on parental antecedents of infant attachment. *Child Development, 68*, 571–591. doi:10.1111/j.1467-8624.1997.tb04218.x

Dykas, M. J., & Cassidy, J. (2011). Attachment and the processing of social information across the life span: Theory and evidence. *Psychological Bulletin, 137*, 19–46. doi:10.1037/a0021367

Fearon, R. P., Bakermans-Kranenburg, M. J., van IJzendoorn, M. H., Lapsley, A., & Roisman, G. I. (2010). The significance of insecure attachment and disorganization in the development of children's externalizing behavior: A meta-analytic study. *Child Development, 81*, 435–456. doi:10.1111/j.1467-8624.2009.01405.x

Fearon, R. P., van IJzendoorn, M. H., Fonagy, P., Bakermans-Kranenburg, M. J., Schuengel, C., & Bokhorst, C. L. (2006). In search of shared and nonshared environmental factors in security of attachment: A behavior-genetic study of the association between sensitivity and attachment security. *Developmental Psychology, 42*, 1026–1040. doi:10.1037/0012-1649.42.6.1026

Flom, R., & Bahrick, L. E. (2007). The development of infant discrimination of affect in multimodal and unimodal stimulation: The role of intersensory redundancy. *Developmental Psychology, 43*, 238–252. doi:10.1037/0012-1649.43.1.238

Forssman, L., Peltola, M. J., Yrttiaho, S., Puura, K., Mononen, N., Lehtimäki, T., & Leppänen, J. M. (2014). Regulatory variant of the TPH2 gene and early life stress are associated with heightened attention to social signals of fear in infants. *Journal of Child Psychology and Psychiatry, 55*, 793–801. doi:10.1111/jcpp.12181

Gergely, G. (2004). The role of contingency detection in early affect-regulative interactions and in the development of different types of infant attachment. *Social Development, 13,* 468–478. doi:10.1111/j.1467-9507.2004.00277.x

Groh, A. M., Fearon, R. P., Bakermans-Kranenburg, M. J., van IJzendoorn, M. H., Steele, R. D., & Roisman, G. I. (2014). The significance of attachment security for children's social competence with peers: A meta-analytic study. *Attachment & Human Development, 16,* 103–136. doi:10.1080/14616734.2014.883636

Groh, A. M., Roisman, G. I., van IJzendoorn, M. H., Bakermans-Kranenburg, M. J., & Fearon, R. P. (2012). The significance of insecure and disorganized attachment for children's internalizing symptoms: A meta-analytic study. *Child Development, 83,* 591–610. doi:10.1111/j.1467-8624.2011.01711.x

Grossmann, T., Striano, T., & Friederici, A. D. (2007). Developmental changes in infants' processing of happy and angry facial expressions: A neurobehavioral study. *Brain and Cognition, 64,* 30–41. doi:10.1016/j.bandc.2006.10.002

Kagan, J., & Herschkowitz, N. (2005). *A young mind in a growing brain.* Mahwah, NJ: Erlbaum.

Leerkes, E. M. (2011). Maternal sensitivity during distressing tasks: A unique predictor of attachment security. *Infant Behavior & Development, 34,* 443–446. doi:10.1016/j.infbeh.2011.04.006

Leppänen, J. M., Forssman, L., Kaatiala, J., Yrttiaho, S., & Wass, S. V. (2014). Widely applicable MATLAB routines for automated analysis of saccadic reaction times. *Behavior Research Methods.* Advance online publication. doi:10.3758/s13428-014-0473-z

Leppänen, J. M., Moulson, M. C., Vogel-Farley, V. K., & Nelson, C. A. (2007). An ERP study of emotional face processing in the adult and infant brain. *Child Development, 78,* 232–245. doi:10.1111/j.1467-8624.2007.00994.x

Leppänen, J. M., & Nelson, C. A. (2009). Tuning the developing brain to social signals of emotions. *Nature Reviews Neuroscience, 10,* 37–47. doi:10.1038/nrn2554

Leppänen, J. M., & Nelson, C. A. (2012). Early development of fear processing. *Current Directions in Psychological Science, 21,* 200–204. doi:10.1177/0963721411435841

Leppänen, J. M., Peltola, M. J., Mäntymaa, M., Koivuluoma, M., Salminen, A., & Puura, K. (2010). Cardiac and behavioral evidence for emotional influences on attention in 7-month-old infants. *International Journal of Behavioral Development, 34,* 547–553. doi:10.1177/0165025410365804

Leppänen, J. M., Peltola, M. J., Puura, K., Mäntymaa, M., Mononen, N., & Lehtimäki, T. (2011). Serotonin and early cognitive development: Variation in the tryptophan hydroxylase 2 gene is associated with visual attention in 7-month-old infants. *Journal of Child Psychology and Psychiatry, 52,* 1144–1152. doi:10.1111/j.1469-7610.2011.02391.x

Luijk, M. P., Roisman, G. I., Haltigan, J. D., Tiemeier, H., Booth-Laforce, C., van IJzendoorn, M. H., & Bakermans-Kranenburg, M. J. (2011). Dopaminergic, serotonergic, and oxytonergic candidate genes associated with infant attachment security and disorganization? In search of main and interaction effects. *Journal of Child Psychology and Psychiatry, 52,* 1295–1307. doi:10.1111/j.1469-7610.2011.02440.x

Main, M., & Hesse, E. (1990). Parents' unresolved traumatic experiences are related to infant disorganized attachment status: Is frightened and/or frightening parental behavior the linking mechanism? In M. T. Greenberg, D. Cicchetti, & E. M. Cummings (Eds.), *Attachment in the preschool years: Theory, research, and intervention* (pp. 161–182). Chicago, IL: University of Chicago Press.

Main, M., & Solomon, J. (1990). Procedures for identifying infants as disorganized/disoriented during the ainsworth strange situation. In M. T. Greenberg, D. Cicchetti, & E. M. Cummings (Eds.), *Attachment in the preschool years: Theory, research and intervention* (pp. 121–160). Chicago, IL: University of Chicago Press.

Marsh, A. A., & Blair, R. J. R. (2008). Deficits in facial affect recognition among antisocial populations: A meta-analysis. *Neuroscience & Biobehavioral Reviews, 32,* 454–465. doi:10.1016/j.neubiorev.2007.08.003

Marsh, A. A., Kozak, M. N., & Ambady, N. (2007). Accurate identification of fear facial expressions predicts prosocial behavior. *Emotion, 7,* 239–251. doi:10.1037/1528-3542.7.2.239

Merras-Salmio, L., Salo, S., Pelkonen, A. S., Kuitunen, M., Aronen, E. T., Mäkelä, M. J., & Kolho, K.-L. (2013). How mothers interact with children with suspected cow's milk allergy symptoms. *Acta Paediatrica, 102,* 1180–1185. doi:10.1111/apa.12408

Mesman, J., van IJzendoorn, M. H., & Bakermans-Kranenburg, M. J. (2009). The many faces of the still-face paradigm: A review and meta-analysis. *Developmental Review, 29,* 120–162. doi:10.1016/j.dr.2009.02.001

Nakagawa, A., & Sukigara, M. (2012). Difficulty in disengaging from threat and temperamental negative affectivity in early life: A longitudinal study of infants aged 12 to 36 months. *Behavioral and Brain Functions, 8,* 40. doi:10.1186/1744-9081-8-40

Nelson, C. A., & de Haan, M. (1996). Neural correlates of infants' visual responsiveness to facial expressions of emotion. *Developmental Psychobiology, 29,* 577–595. doi:10.1002/(SICI)1098-2302(199611)29:7 < 577::AID-DEV3 > 3.0.CO;2-R

Nelson, C. A., & Dolgin, K. G. (1985). The generalized discrimination of facial expressions by seven-month-old infants. *Child Development, 56,* 58–61.

Olson, K. R., & Dweck, C. S. (2009). Social cognitive development: A new look. *Child Development Perspectives, 3,* 60–65. doi:10.1111/j.1750-8606.2008.00078.x

Peltola, M. J., Hietanen, J. K., Forssman, L., & Leppänen, J. M. (2013). The emergence and stability of the attentional bias to fearful faces in infancy. *Infancy, 18,* 905–926. doi:10.1111/infa.12013

Peltola, M. J., Leppänen, J. M., & Hietanen, J. K. (2011). Enhanced cardiac and attentional responding to fearful faces in 7-month-old infants. *Psychophysiology*, *48*, 1291–1298. doi:10.1111/j.1469-8986.2011.01188.x

Peltola, M. J., Leppänen, J. M., Mäki, S., & Hietanen, J. K. (2009). Emergence of enhanced attention to fearful faces between 5 and 7 months of age. *Social Cognitive and Affective Neuroscience*, *4*, 134–142. doi:10.1093/scan/nsn046

Peltola, M. J., Leppänen, J. M., Palokangas, T., & Hietanen, J. K. (2008). Fearful faces modulate looking duration and attention disengagement in 7-month-old infants. *Developmental Science*, *11*, 60–68. doi:10.1111/j.1467-7687.2007.00659.x

Peltola, M. J., Leppänen, J. M., Vogel-Farley, V. K., Hietanen, J. K., & Nelson, C. A. (2009). Fearful faces but not fearful eyes alone delay attention disengagement in 7-month-old infants. *Emotion*, *9*, 560–565. doi:10.1037/a0015806

Pollak, S. D., & Kistler, D. J. (2002). Early experience is associated with the development of categorical representations for facial expressions of emotion. *Proceedings of the National Academy of Sciences of the United States of America*, *99*, 9072–9076. doi:10.1073/pnas.142165999

Rothbart, M. K. (1981). Measurement of temperament in infancy. *Child Development*, *52*, 569–578.

Schuengel, C., Bakermans-Kranenburg, M. J., & van IJzendoorn, M. H. (1999). Frightening maternal behavior linking unresolved loss and disorganized infant attachment. *Journal of Consulting and Clinical Psychology*, *67*, 54–63. doi:10.1037/0022-006X.67.1.54

Soken, N. H., & Pick, A. D. (1999). Infants' perception of dynamic affective expressions: Do infants distinguish specific expressions? *Child Development*, *70*, 1275–1282. doi:10.1111/1467-8624.00093

Sroufe, L. A. (2005). Attachment and development: A prospective, longitudinal study from birth to adulthood. *Attachment & Human Development*, *7*, 349–367. doi:10.1080/14616730500365928

Tharner, A., Herba, C. M., Luijk, P. C. M., van IJzendoorn, M. H., Bakermans-Kranenburg, M. J., Govaert, P. P., . . . Tiemeier, H. (2011). Subcortical structures and the neurobiology of infant attachment disorganization: A longitudinal ultrasound imaging study. *Social Neuroscience*, *6*, 336–347. doi:10.1080/17470919.2010.538219

Van IJzendoorn, M. H., & Bakermans-Kranenburg, M. J. (2012). Integrating temperament and attachment: The differential susceptibility paradigm. In M. Zentner & R. L. Shiner (Eds.), *Handbook of temperament* (pp. 403–424). New York, NY: Guilford.

Van IJzendoorn, M. H., Schuengel, C., & Bakermans-Kranenburg, M. J. (1999). Disorganized attachment in early childhood: Meta-analysis of precursors, concomitants, and sequelae. *Development and Psychopathology*, *11*, 225–249.

Vrtička, P., & Vuilleumier, P. (2012). Neuroscience of human social interactions and adult attachment style. *Frontiers in Human Neuroscience*, *6*, 212. doi:10.3389/fnhum.2012.00212

Walle, E. A., & Campos, J. J. (2012). Interpersonal responding to discrete emotions: A functionalist approach to the development of affect specificity. *Emotion Review*, *4*, 413–422. doi:10.1177/1754073912445812

Yrttiaho, S., Forssman, L., Kaatiala, J., & Leppänen, J. M. (2014). Developmental precursors of social brain networks: The emergence of attentional and cortical sensitivity to facial expressions in 5 to 7 months old infants. *PLoS ONE*, *9*, e100811. doi:10.1371/journal.pone.0100811

Zieber, N., Kangas, A., Hock, A., & Bhatt, R. S. (2014). Infants' perception of emotion from body movements. *Child Development*, *85*, 675–684. doi:10.1111/cdev.12134

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's website:

Data S1. Analyses of the 5-Month Data

Child Development, September/October 2015, Volume 86, Number 5, Pages 1333–1350

Temperament and Interparental Conflict: The Role of Negative Emotionality in Predicting Child Behavioral Problems

Rochelle F. Hentges and Patrick T. Davies
University of Rochester

Dante Cicchetti
Mt. Hope Family Center, University of Rochester and University of Minnesota

This study examined temperamental irritability and fearful distress as moderators of the association between interparental conflict and child behavior problems in a disadvantaged sample of two hundred and one 2-year-old children and their mothers. Using a multimethod, prospective design, findings revealed that the relation between interparental conflict and changes in child behavior problems over a 1-year period were moderated by temperamental irritability. Consistent with differential susceptibility theory, children high in irritable temperament not only exhibited poorer outcomes in contexts of high interparental conflict but also better adjustment in contexts of low levels of interparental conflict. Mediated moderation analyses revealed that fearful reactivity partly accounted for the greater susceptibility of irritable children, particularly in explaining why they fared better when interparental conflict was low.

Exposure to high levels of interparental conflict has been shown to increase children's risk for developing disruptive behavior problems (Crockenberg & Langrock, 2001). However, the modest to moderate effect sizes identified in research also highlight the considerable heterogeneity in children's adjustment to interparental conflict (Fincham, 1994). Toward the goal of identifying sources of variability in outcomes of children who witness similar levels of interparental conflict, conceptual models point to the potential moderating role of temperament (e.g., Crockenberg & Langrock, 2001; Davies & Cummings, 1994; Grych & Fincham, 1990). Although empirical research has lagged behind theoretical advances, these models share the assumption that susceptibility to negative emotional distress is a key mechanism that may account for these individual differences in adjustment. Negative emotionality, which is defined by the disposition to experience negative emotions such as anger and fear, is presumed to increase the risk for developing psychological problems in adverse contexts (Rothbart, 2011). Accordingly, a primary objective of this article is to identify whether children's temperamental disposition to experience specific forms of negative emotionality may moderate associations between interparental conflict and their risk for disruptive behavior problems. To better understand the mechanisms that account for the moderating role of negative emotionality, we further examine whether children's emotional responses to interparental conflict mediate the interaction between interparental conflict and negative emotionality in predicting children's behavior problems.

As a framework for organizing our aims, Figure 1 provides a conceptual model of the hypothesized interplay between negative emotionality and interparental conflict predicting child disruptive behavior problems. As depicted by Path 1 in Figure 1, we hypothesize that temperamental negative emotionality moderates the pathway between interparental conflict and children's behavior problems. Guided by conceptualizations of temperamental traits as having constitutional origins, emerging early in life, and remaining relatively resistant to change as a function of environmental conditions (Rothbart, 2011), multiple theories have proposed that temperamental negative emotionality is a moderator of interparental conflict (Crockenberg & Langrock, 2001; Cummings & Davies, 2010; Grych & Fincham, 1990). However, there is little empirical research on the interplay between temperamental

---

This research was supported by the National Institute of Mental Health (R01 MH071256) awarded to Patrick T. Davies and Dante Cicchetti. The project was conducted at Mt. Hope Family Center. The authors are grateful to the children, parents, and community agencies who participated in this project and to the Mt. Hope Family Center staff.

Correspondence concerning this article should be addressed to Rochelle F. Hentges, Department of Clinical and Social Sciences in Psychology, University of Rochester, Rochester, NY 14627. Electronic mail may be sent to rhentges@ur.rochester.edu.

© 2015 The Authors
Child Development © 2015 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2015/8605-0003
DOI: 10.1111/cdev.12389

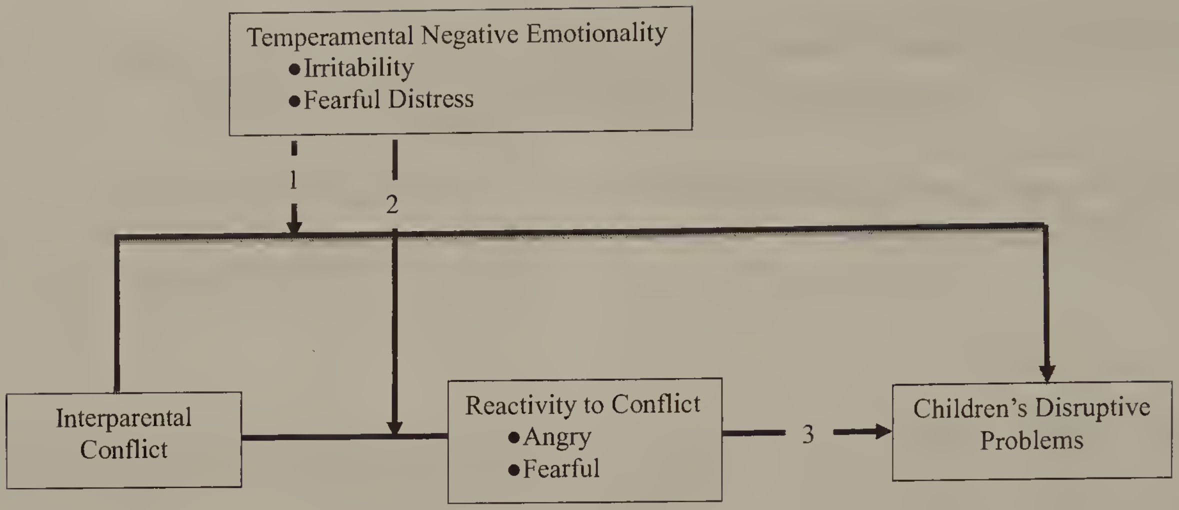
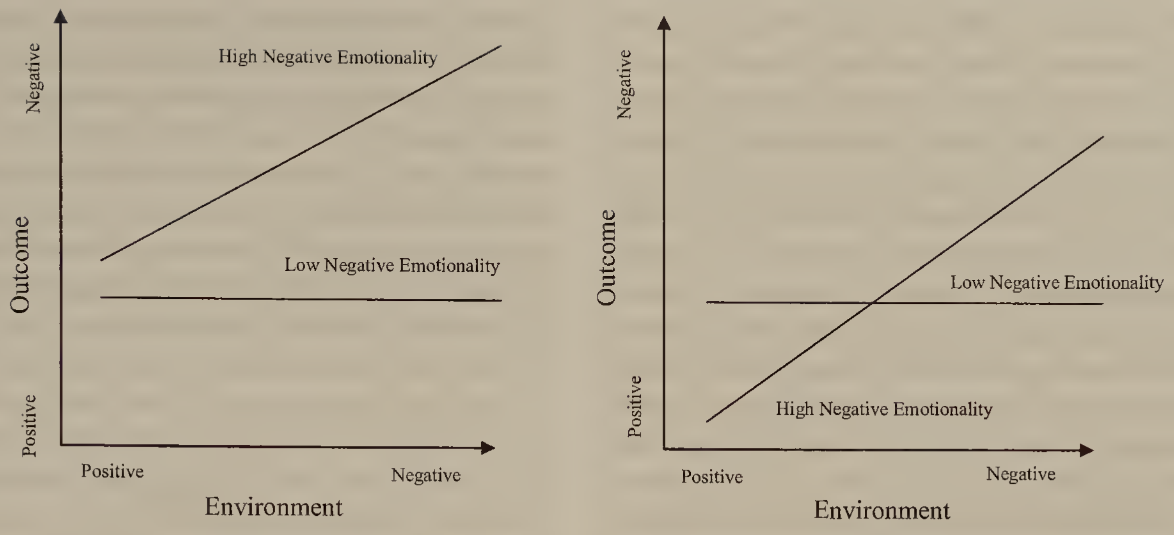


Figure 1. A conceptual model of the proposed mediated moderation framework in which negative emotionality moderates the link between interparental conflict and child disruptive behavior problems through the mechanism of child reactivity to interparental conflict.

negative emotionality and interparental conflict. As a notable exception, Pauli-Pott and Beckmann (2007) reported that interparental conflict predicted behavior problems 2 years later for infants who exhibited high, but not low, negative emotionality. Our study is designed to build on these promising findings in several ways.

First, we extend this research direction into the early preschool years as this developmental period may be particularly significant for understanding individual differences in children's psychological difficulties. Not only do preschool children appear to experience more frequent interparental conflict than older children, but research also suggests that they exhibit greater sensitivity to conflict between parents than younger children (Fantuzzo, Boruch, Beriama, Atkins, & Marcus, 1997; Kitzmann, Gaylord, Holt, & Kenny, 2003). Disruptive behavior problems characterized by oppositional defiant disorder (ODD) and attention deficit hyperactivity disorder (ADHD) symptoms are particularly prevalent as sequelae of family conflict during the preschool period (Cunningham & Boyle, 2002; Goldstein et al., 2007). Evidence indicates that early childhood is a sensitive period for the emergence of organized patterns of these forms of psychopathology (Shaw, Bell, & Gilliom, 2000), which in turn increase children's risk for experiencing a wide array of adjustment difficulties in social, emotional, academic, and physical health domains (Cunningham & Boyle, 2002). Accordingly, our focus in this article is on examining individual differences in children's ODD and ADHD symptoms in contexts of interparental conflict.

Second, in the context of new conceptual developments on different forms of moderating effects of temperament, the Pauli-Pott and Beckmann (2007) findings were not designed to definitively test the relative viability of two prevailing organismic-environment models. On the one hand, diathesis-stress models, as shown in Figure 2a, conceptualize temperamental negative emotionality as a diathesis or a risk factor that emerges as particularly potent in the context of environmental adversity (Rothbart, 2011). Therefore, whereas negative emotionality may be more modestly or even negligibly associated with disruptive problems when destructive (e.g., escalating hostility, aggression) conflict is low, diathesis-stress models posit that negative emotionality results in substantially greater vulnerability for psychopathology under high levels of destructive interparental conflict (e.g., Colder, Lochman, & Wells, 1997; Morris et al., 2002). On the other hand, differential susceptibility theory proffers that negative emotionality serves as a plasticity factor taking a disproportionate toll on children's functioning in the face of environmental adversity but potentially conferring significant advantages within highly supportive contexts (Belsky & Pluess, 2009). Figure 2b specifically shows that diathesis-stress and differential susceptibility models share the hypothesis that children with high negative emotionality will fare significantly worse than children with low negative emotionality in discordant environments. Although alternative diathesis-stress models have been proposed (see Ingram & Luxton, 2005), the key source of distinction between diathesis-stress and differential susceptibility models is that differential susceptibility theory postulates that children higher in negative emotionality will fare significantly better under supportive socialization conditions. Although a growing corpus of studies have found support for


Negative
Outcome
Positive
High Negative Emotionality
Low Negative Emotionality
Positive
Negative
Environment
Negative
Outcome
Positive
Low Negative Emotionality
High Negative Emotionality
Positive
Negative
Environment

(a) Diathesis-Stress Model

(b) Differential Susceptibility Model

Figure 2. Diathesis-stress (a) versus differential susceptibility (b) models of Temperament × Environment interaction.

negative emotionality as a susceptibility or plasticity factor in models of parenting (e.g., Lengua, 2008; Poehlmann et al., 2011), the comparative utility of diathesis-stress and differential susceptibility frameworks have yet to be tested in models of interparental conflict and child temperament.

Third, as denoted in Path 1 of Figure 1, our objective of understanding the interplay between interparental conflict and temperament in early childhood highlights the potential value of distinguishing between different forms of negative emotionality. As children progress through early childhood, their behavioral expressions of negative emotionality become increasingly differentiated into dimensions of negative emotionality characterized by: (a) irritability and anger, and (b) vulnerable forms of distress manifested in fear and sadness (Lemery, Goldsmith, Klinnert, & Mrazek, 1999; Rothbart, 2011). Each of these dimensions of negative emotionality is specifically proposed to reflect different neurobiological systems for regulating organismic interactions in challenging contexts (Buss & Goldsmith, 1998). Anger and irritability are functionally proposed to be products of an approach-oriented system that is organized around dominance and the removal of challenges and threats (Derryberry & Rothbart, 1997; Waters & Thompson, 2014). Conversely, fear and sadness are theorized to be manifestations of a withdrawal system that is designed to engender submissive ways of coping with environmental stressors (Korte, Koolhaas, Wingfield, & McEwen, 2005; Waters & Thompson, 2014). Thus, in this study, we examine whether temperamental irritability and fearful distress operate as potentially distinct moderators in associations between interparental conflict and disruptive behavior problems.

Finally, documenting the moderating effects of temperament is only a first step toward understanding the heterogeneity of outcomes of children exposed to interparental conflict. Precisely delineating the form of moderation for temperamental irritability and fearful distress can be facilitated through an understanding of why associations between interparental conflict and children's disruptive problems may vary across the two dimensions of negative emotionality. Thus, a key remaining task is to identify mechanisms that are more proximal products of the interplay between interparental conflict and negative emotionality and, in turn, act as precursors to trait-like patterns of disruptive behavior problems. Theories of interparental conflict have consistently proposed that children's angry and fearful emotional reactivity to interparental conflict are key mechanisms mediating the moderating role of temperament in associations between interparental conflict and children's psychological maladjustment (e.g., Crockenberg & Langrock, 2001; Cummings & Davies, 2010). This hypothesis is rooted in the conceptualization of forms of emotional reactivity as relational constructs that: (a) are highly sensitive to the confluence of children's temperamental characteristic and prior experiences with interparental conflict, and (b) serve as precursors of subsequent psychological problems (Cummings & Davies, 2010).

In accordance with these assumptions, emotional security theory has postulated that interparental conflict differentially predicts children's fearful reactivity to interparental conflict depending on children's preexisting tendencies to experience negative emotionality. As a sign of emotional insecurity, fearful reactivity to interparental conflict, in turn, is further theorized to be an operative mechanism underlying children's vulnerability to disruptive behavior problems. Specific emotions theory (Crockenberg & Langrock, 2001) and social learning theories of interparental conflict (e.g., Emery, 1989) also call attention to the role of angry reactivity to conflict as a potential mediator in the link between interparental conflict and child behavior problems. Thus, we specifically examine a mediated moderation model in which children's angry and fearful emotional reactions to interparental conflict account for why forms of negative emotionality moderate the path between interparental conflict and children's disruptive behavior problems.

Process models of interparental conflict also provide some clues into how different forms of negative emotionality may alter the risk posed by interparental conflict. For clarity, our formulation of the mediated moderation is conceptually addressed in a two-step process corresponding to each link in the mediational chain. The first link outlined in Path 2 of Figure 1 highlights the possibility that children may experience different degrees of sensitization to interparental conflict depending on preexisting levels of their temperamental irritability and fearful distress. An analysis of differences in the emotional processing and coping capacities of temperamentally irritable and fearful children raises several possible predictions about how associations between interparental conflict and children's negative emotional reactivity to interparental conflict may differ as a function of each form of negative emotionality.

Temperamental irritability has been linked with diminished emotion regulation and coping abilities (Blair, Denham, Kochanoff, & Whipple, 2004; Calkins, Dedmon, Gill, Lomax, & Johnson, 2002). Thus, when repeatedly exposed to high levels of destructive interparental conflict, the limited coping capacities of highly irritable children may be manifested in more intense and prolonged expressions of emotional reactivity to interparental conflict. Alternatively, high irritability may actually dampen children's expressions of vulnerable (e.g., fearful) emotional reactivity to interpersonal challenges, particularly in settings that are not laden with imminent threats (e.g., Korte et al., 2005; Waters & Thompson, 2014). Thus, children high in irritability may be able to more easily inhibit expressions of fear in response to interparental conflict, at least when their prior histories of experiences with interparental conflict are relatively benign. Temperamental dispositions to experience fearful distress, on the other hand, have been linked with increased vigilance and sensitivity to cues of interpersonal threat (LeDoux, 2000). Therefore, children high in fearful distress may exhibit higher levels of fearful reactivity to interparental conflict when they are exposed to heightened destructive interparental conflict.

As the final link of the mediated moderation chain, Path 3 in Figure 1 illustrates the hypothesis that angry and fearful reactivity resulting from the interplay between interparental conflict and temperamental negative emotionality increase children's vulnerability to disruptive behavior problems. According to emotional security theory, intense and prolonged expressions of fearful reactivity to interparental conflict tip the balanced allocation of psychobiological resources toward investing in immediate personal safety in the family at the cost of investment in resources for mastery of other important goal pursuits. Resulting deficits in self-regulatory abilities may, over time, increase oppositional and defiant behavior, hyperactivity, impulsivity, and inattention (Davies & Sturge-Apple, 2007). Likewise, several conceptual models have highlighted the significance of angry reactivity to interparental conflict as a prognosticator of disruptive problems though its tendency to engender callousness and hostile world views (Crockenberg & Langrock, 2001; Davies & Sturge-Apple, 2007).

In summary, the current study is designed to empirically test the conceptual role of negative emotionality as a moderator in the association between interparental conflict and preschool children's disruptive behavior problems. In adopting a trait-specific approach, a key aim is to distinguish between two forms of negative emotionality (i.e., irritability and fearful distress) as a way to account for the heterogeneity in the adjustment of children exposed to interparental conflict. As part of this goal, we specifically examine the relative viability of diathesis-stress and differential susceptibility models in characterizing the nature of the moderating effects of negative emotionality. In breaking further ground, a final aim of this article is to move beyond simply cataloguing moderators of interparental conflict by exploring possible mediating mechanisms. Utilizing a mediated moderation approach within a multimethod longitudinal design, we test children's fearful and angry responses to interparental conflicts as mediators of any moderating effects of temperamental emotionality. Because previous studies have relied heavily on predominantly White, middle-class samples (e.g., Davies & Windle, 2001), we build on a growing corpus of research on the significance of interparental conflict for children from different racial backgrounds by examining these research questions in a racially diverse sample of families experiencing significant impoverishment (McLoyd, Harper, & Copeland, 2001; Spjeldnes & Choi, 2008).

Method

Participants

Participants included two hundred and one 2-year-old children and their mothers from a Northeastern metropolitan area between 2005 and 2008 as part of a larger project designed to explore the impact of interparental conflict on children. Our recruitment aim was to obtain a racially diverse sample of families experiencing a wide range of interparental conflict that simultaneously limited the confounding role of socioeconomic characteristics (Fergusson & Horwood, 1998; Kitzmann et al., 2003). Therefore, in the first step, we recruited participants through agencies serving disadvantaged children and families including Women, Infants, and Children; Temporary Assistance to Needy Families rosters from the Department of Human and Health Services; and the county family court system. In the second step, we administered the abbreviated version of the Physical Assault Subscale of the Conflict Tactics Scale 2 (CTS2; Straus, Hamby, Boney-McCoy, & Sugarman, 1996) to insure that roughly equal proportions of participating mothers fell within three bands of an interparental conflict continuum: (a) no violence (40%), (b) mild/moderate physical violence (24%), and (c) severe physical violence (36%) in the interpartner relationship. Additional inclusionary criteria were implemented to ensure that the female caregiver was the biological mother and the male partner had regular contact with the mother and toddler over the past year. In addition, all participants were required to speak and read English to ensure all measures could be adequately administered and coded.

Median annual income for the family household was $18,300 (U.S.) per year and a substantial minority of mothers (30%) and their partners (24%) did not complete high school. Most families received public assistance (95%) and were impoverished according to the U.S. Federal Poverty Guidelines (99.5%). The mean age of the children was 26 months ($SD = 1.69$), with 44% of the sample consisting of girls ($n = 92$). Mothers ranged in age from 18 to 44 at the first wave ($M = 26.24$). The majority of mothers and children were Black (56%), followed by smaller proportions of family members who identified as White (23%), Latino (11%), Multi-Racial (7%), and "Other" (3%). The retention rate across the two measurement occasions was 83%. To test for selective attrition, statistical comparisons were conducted between the mother–child dyads who participated in both occasions and the dyads who did not participate in Wave 2 across the primary variables, covariates, and demographic characteristics at Wave 1 (e.g., interparental conflict, income). Families who were retained at both waves did not significantly differ from those who dropped out on any of the variables.

Procedure

Mothers and children participated in multiple visits to a research center laboratory at each of two waves beginning when children were 2 years old. The two measurement occasions were spaced 1 year apart. All procedures were approved by the Institutional Review Board at the research site prior to conducting the study.

Interparental Disagreement Interview and Questionnaires

During Wave 1, mothers completed a number of questionnaires to assess the frequency and nature of interparental conflict and participated in the Interparental Disagreement Interview (IDI). As a well-established measure (e.g., Davies, Cicchetti, & Martin, 2012), the IDI contains two modules that are designed to capture: (a) interparental conflict characteristics, which enable the assessment of the frequency, course, and aftermath of common interparental disagreements, and (b) child reactivity to conflict, which captures children's emotional and behavioral responses to interparental disagreements. After selecting a topic of interparental disagreement that commonly takes place in front of the child, a trained experimenter presented mothers with a series of open-ended questions within each module. Selective examples of questions included in the interview are "How would you describe your disagreements over [topic]?" "What do you/your partner say or do when you disagree?" and "During these disagreements that [child name] sees or hears, how does s/he respond?" Additional probes are implemented when necessary to maximize the richness of maternal narratives.

Unfamiliar Episodes

During the first wave of the study, the toddlers were exposed to a series of unfamiliar and challenging episodes designed to assess temperamental differences in reactivity, including fearful and irritable emotionality. The procedures were comparable to previous temperament batteries (e.g., Braungart-Rieker, Hill-Soderlund, & Karrass, 2010; Fox, Henderson, Rubin, Calkins, & Schmidt, 2001; Kagan, Reznick, & Snidman, 1987). Mothers sat in a corner of the room but were instructed to intervene only if they became concerned about their child's well-being. During the first episode, children were free to explore a number of unusual objects in the room (e.g., windshield cover). In the second episode, an experimenter entered the room and instructed the child to manipulate each object in a specific way (e.g., funnel: "Put it on your head!"; windshield cover: "Poke it!"). For the third episode, an unfamiliar female experimenter dressed as a clown invited the toddlers to play with a bag of toys after introducing herself and then left with the bag of toys after 5 min. In the fourth episode, the initial experimenter reentered the room with a large robot holding an attractive toy on a tray. Operated surreptitiously by remote control from the control room, the robot alternated three times between designated periods of inactivity and movement around the room. Each time the robot stopped moving, the experimenter encouraged the child to approach the toy. The final episode involved the primary experimenter asking the child to imitate a series of events after first enacting them herself, including (a) reaching behind a black curtain to pull out a doll, (b) placing a finger in glasses of water and prune juice, and (c) picking up a rubber snake and letting it slide back onto the table. These tasks were designed to be challenging and thus elicit individual differences in emotionality, including fear to novel objects, people, and events as well as irritable reactivity to the challenge of performing within the constraints of the tasks (e.g., playing with the toys in a specified manner, watching the clown depart with fun toys). Video records of the unfamiliar episodes were subsequently coded for temperamental irritability and fearful distress.

Child Behavior Problems

At both waves, mothers completed a questionnaire to assess children's behavior problems.

Measures

Interparental Conflict

Definitive tests of the relative power of diathesis-stress and differential susceptibility predictions require assessment of a broad spectrum of environmental conditions ranging from supportive to adverse. Therefore, in line with recommendations for testing for differential susceptibility (Belsky & Pluess, 2009), survey and interview measures of both constructive and destructive forms of interparental conflict were included in a composite measure of interparental conflict. For the first two indices of destructive conflict, mothers completed the Physical Aggression Subscale of the Conflict and Problem-Solving (CPS) scales (Kerig, 1996) and the Psychological Aggression Subscale of the Conflict Tactics Scale–Revised (CTS2; Straus et al., 1996). The CPS Physical Aggression subscale contains 14 items that assess the extent to which mothers and their partners engage in a number of physically aggressive behaviors during conflicts. The CTS2 Psychological Aggression subscale consists of 16 items designed to assess occurrences of maternal and partner psychological abusive acts in the interparental relationship in the past year. Consistent with recommendations (Straus et al., 1996), each item is quantified as 0 (*no occurrence*) or 1 (*one or more occurrences*). The CPS Physical Aggression and CTS2 Psychological Aggression scales evidenced satisfactory reliability (αs = .92 and .88, respectively) and prior research supports the validity of each scale (Kerig, 1996; Straus et al., 1996). As the final indicator of destructive conflict, trained observers rated aggression based on maternal narratives from the interparental conflict characteristics module of the IDI. The aggression code is rated on a scale of 0–6 (*none–high*), reflecting the degree to which partners engaged in aggressive acts during the disagreements being discussed during the interview. High aggression (i.e., 6) reflected considerable dysregulation, disorganization, or loss of control by the parent, which reflects a clear risk to the psychological or physical welfare of the child. Intraclass correlation coefficients, indexing reliability between two raters who independently coded 26% of the videotapes, were .81 and .78 for maternal and partner aggression, respectively. Due to their high correlation (r = .54), mother and partner aggression ratings were averaged to form a single measure of interpartner aggression.

Two measures served as indicators of constructive conflict. First, as a maternal survey assessment, the CPS Resolution subscale contains 13 items assessing the extent to which the couple attempts to resolve conflicts together (e.g., “Try to find a solution that meets both of our needs equally”). Reliability for the CPS Resolution scale was good ($\alpha = .87$) and its validity is supported by prior research (e.g., Kerig, 1996). Second, coders also rated the interparental conflict characteristics module of the IDI for supportive problem solving. ranging from 0 (*none*) to 6 (*high*); supportive problem solving indexes the degree to which each parent engaged in constructive problem solving during disagreements (e.g., identifying the problem, engaging in perspective taking, generating constructive solutions to the problem). Single-measure intraclass correlation coefficients, indexing reliability between two coders who independently rated 26% of the videotapes, were .64 for both the mom and partner. Maternal and partner supportive problem-solving ratings were aggregated together to form an interparental problem-solving composite due to their high correlation ($r = .81$). To obtain an overall measure of children’s experiences with both constructive and destructive interparental conflict, the five measures of interpartner conflict were standardized after reverse scoring the constructive conflict scales. Internal consistency for the five-indicator composite was $\alpha = .71$.

Temperament

Consistent with global rating approaches capturing different forms of negative emotionality (Durbin, Klein, Hayden, Buckley, & Moerk, 2005; Hayden, Klein, Durbin, & Olino, 2006), two trained, independent observers coded the novel task procedures for temperamental irritability and fearful distress on 5-point scales (0 = *none*, 4 = *intense*) during each of the five episodes, overlapping on 25% of the observations. The trained observers were instructed to pay attention to facial, vocal, and body signals of fearful distress and irritability throughout each task and assign one global code for each episode, resulting in a total of five global codes. Temperamental irritability was defined by signs of anger, frustration, and hostility, such as pushing the robot or angrily refusing to comply with instructions during the imitation tasks. Fearful distress was evidenced by signs of fear, worry, and dysphoria, such as whining, crying, fidgeting, or comfort seeking. Observer ratings across the five episodes were averaged together to create composites of temperamental irritability and fearful distress. Mean levels of irritability ($M = 1.23$, $SD = 0.82$) and fearful distress ($M = 1.33$, $SD = 0.62$) were in the lower middle of the scale and evidenced a normal distribution, suggesting that the tasks adequately captured individual differences in both irritability and fearful distress. Alphas across the five tasks showed adequate internal consistency (.74 for irritability and .58 for fearful distress) and were consistent with alphas obtained using similar temperament batteries to assess fearful and irritable emotionality (Durbin et al., 2005; Van der Mark, Bakermans-Kranenburg, & van IJzendoorn, 2002). Interrater reliability between observers was satisfactory, with intraclass correlation coefficients ranging from .75 to .97 for irritability and fearful distress codes across the five episodes.

Child Behavior Problems

Child behavior problems were assessed at both waves by the Child Behavior Checklist (CBCL/ages 1½–5; Achenbach & Rescorla, 2006) *Diagnostic Statistical Manual of Mental Disorders* (*DSM*)-oriented scales for Attention Deficit/Hyperactivity Problems and Oppositional Defiant Problems. *DSM*-oriented scales were designed to more accurately index forms of psychopathology based on previous evidence that the CBCL syndrome scales exhibit weak relations with childhood disorders as defined by the *DSM* (Nakamura, Ebesutani, Bernstein, & Chorpita, 2009). Consistent with this goal, research has not only supported the reliability and validity of the *DSM*-oriented scales, but also its abilities to more accurately predict disruptive behavioral problems than the CBCL syndrome scales (Aebi, Metzke, & Steinhausen, 2010; Bellina et al., 2013). Thus, we opted to capture disruptive child behavior problems with the *DSM*-oriented scales. The Attention Deficit/Hyperactivity Scale is designed to assess hyperactivity and inattention (e.g., “can’t sit still, restless or hyperactive”), whereas the Oppositional Defiant Subscale captures defiance and outward displays of anger (e.g., “disobedient,” “temper tantrums or hot temper”). Internal consistencies ranged from .74 to .85 across the two time points.

Child Reactivity to Interparental Conflict

Video records of the child reactivity module of the IDI were subsequently coded by a pair of trained coders who were different from the observers for the interparental conflict characteristics component of the IDI. Children’s fearful and angry reactivity to interparental conflict constructs were each assessed through three codes. For the first indicator of each form of reactivity, coders assessed the presence (1) or absence (0) of children's specific responses to interparental conflict at each of two time periods: during the conflict and immediately after the conflict. The specific behaviors reflecting fearful reactivity included: (a) expression of fear, defined as facial expressions, postural displays, or verbalizations reflecting anxiety or fear; (b) whining, evidenced by whimpering or sobbing; and (c) active flight, reflected in vigorous attempts to avoid or escape the conflict. Target behaviors reflecting angry reactivity included: (a) anger, characterized by descriptions of facial expressions, behaviors (e.g., stomping feet, clenching fists), or verbalizations (e.g., yelling) reflective of anger; (b) aggression, defined by verbal or physical hostility (e.g., hitting, pushing, throwing objects, mocking); and (c) risky or bold involvement in the disagreement (e.g., demanding the parents to stop, taking sides in the disagreement). Specific *behavioral counts* were aggregated during and after the disagreement to form tallies of fearful and angry reactivity, ranging from 0 to 6.

As an additional indicator of each form of reactivity, coders provided specific *continuous ratings* of children's fearful reactivity and hostility along a 7-point scale both during and after the disagreement. The multiplicity, quality, and intensity of emotional responses were considered when assigning the codes. Narratives that contained no indicator of the specific emotion were coded as 0, whereas a code of 6 on the continuous scale reflected children who expressed dysregulated and disturbing patterns of emotional reactivity to the conflict, indicating significant problems controlling and regulating the specific emotion. Codes of each pattern of responding during and after the conflict were averaged to form *continuous ratings* of fearful reactivity and hostility over the two time periods.

For the final pair of indicators, *molar ratings* of child anxious and angry reactivity to conflict were assigned based on the overall narrative and organizational profile of the child's behavioral responses during and after the disagreement. Ratings reflected the degree to which the mothers' narratives corresponded to prototypical profiles of anxious or angry responses to conflict on a 5-point scale, ranging from *not at all characteristic* (1 = *none or minimal signs*) to *highly characteristic* (5 = *telltale signs of the emotional pattern*). High molar ratings of anxious reactivity were characterized by blatant, intense signs of fear in combination with displays of anxious involvement in (e.g., asking questions, moving closer to observe) or avoidance of parental conflict (e.g., fleeing the room). High molar ratings of angry reactivity reflected a high degree of overtly hostile, aggressive, bossy, and demanding behaviors (e.g., pushing one or both parents, demanding one parent leave the other alone). To provide an index of interrater reliability for each measure of conflict reactivity, another coder independently rated 25% of the video records. For the specific behaviors comprising the behavioral tallies, kappa coefficients of interrater reliability ranged from 0.82 to 1.00. Intraclass correlation coefficients for the behavioral tallies, continuous ratings, and molar ratings were all excellent, ranging from .91 to .99.

Covariates

Due to their documented associations with child adjustment, child gender, family income, and maternal education were treated as covariates. Mothers completed a demographic interview at the first wave to obtain indices of annual earned income and maternal education on a scale ranging from 1 (*no formal education to seventh grade*) to 7 (*masters or doctoral degree*).

Results

Table 1 provides the means, standard deviations, and correlations among the variables included in the study. Although raw means are depicted in the table, earned income evidenced significant skewness and was subsequently log transformed to increase normality in the distribution of scores. To retain our full sample size, missing data (Mdn = 8.5%, range = 0%–17.4%) were estimated using full information maximum likelihood in Amos 18.0 (Arbuckle, 2009) for all primary analyses.

Primary Analyses

To examine the moderating role of child temperament within a prospective model, we calculated changes in children's disruptive behavior problems across the two annual measurement occasions through latent difference score (LDS) models using structural equation modeling (SEM; McArdle, 2009). LDS offers a powerful way of testing predictors of interindividual differences in intraindividual changes in behavior problems (McArdle, 2009). In accord with standard LDS procedures, the two components of our dual change model consisted of: (a) a growth parameter reflecting change in

Table 1

Means, Standard Deviations, and Correlations of the Primary Variables in the Analyses

	M	*SD*	Range	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. Earned income	11.21	14.73	0.00 to 90.70	—														
2. Maternal education	4.10	1.42	1.00 to 7.00	.11	—													
3. Temperamental irritability	1.23	0.82	0.00 to 3.80	−.08	−.09	—												
4. Temperamental fear	1.21	0.62	0.00 to 3.60	.06	.07	−.02	—											
5. Interparental conflict	0.00	0.69	−1.64 to 2.58	−.14	.05	−.10	−.00	—										
Child fearful reactivity to conflict																		
6. Beh. count	1.07	0.96	0.00 to 4.00	.00	−.04	−.21*	.03	.25*	—									
7. Cont. rating	1.43	1.34	0.00 to 5.00	−.03	−.03	−.18*	−.01	.32*	.68*	—								
8. Molar rating	2.53	1.30	1.00 to 5.00	−.05	−.03	−.08	.01	.21*	.49*	.60*	—							
Child angry reactivity to conflict																		
9. Beh. count	1.79	1.62	0.00 to 6.00	−.08	−.03	.13	−.07	.05	.09	−.04	.32*	—						
10. Cont. rating	0.84	1.23	0.00 to 5.00	−.12	−.07	.14	−.07	.16*	.06	−.03	.25*	.87*	—					
11. Molar rating	1.74	1.15	1.00 to 5.00	−.19*	.00	.14	−.07	.15*	.03	−.11	.17*	.76*	.84*	—				
Time 1: Child behavior problems																		
12. ADHD	5.60	2.81	0.00 to 12.00	−.10	−.14	.24*	−.07	.06	−.03	−.02	.00	.08	.12	.03	—			
13. ODD	4.09	2.79	0.00 to 12.00	−.04	.05	.17*	.00	.23*	.03	.06	.08	.12	.17*	.12	.63*	—		
Time 2: Child behavior problems																		
14. ADHD	5.17	2.76	0.00 to 12.00	−.16*	−.09	.18*	−.04	.17*	.19*	.13	.11	.15	.20*	.13	.65*	.52*	—	
15. ODD	3.69	2.96	0.00 to 11.00	.03	.06	.11	.01	.29*	.18*	.15	.18*	.11	.14	.11	.37*	.62*	.58*	—

Note. Beh. count = behavioral counts; Cont. rating = continuous ratings, ADHD = attention deficit hyperactivity disorder; ODD = oppositional defiant disorder.

$^{*}p \leq .05$.

behavior problems across the two measurement occasions, and (b) an autoregressive path that estimates the effect of initial behavior problems on behavior problems at the second wave. Thus, the LDS model offers a rigorous way to test for change in levels of behavior problems while controlling for the effect of the initial status on change in problems over time.

Forms of Negative Emotionality as Moderators of Interparental Conflict

Testing whether child emotional responses to interparental conflict partially explains the moderating effects of child temperament first requires demonstrating that forms of negative emotionality significantly interact with interparental conflict in predicting child disruptive behavior problems over time. To examine the first step, each form of negative emotionality (i.e., irritability, fearful distress) was modeled separately in SEM to simultaneously specify the interparental conflict composite, child negative emotionality, and their multiplicative interaction as predictors of the latent intercept and difference factors in the LDS model of child behavior problems. Loadings of each of the indicators of behavior problems were constrained to be equal across the measurement occasions to maximize measurement equivalence. Child gender, household income, and maternal education were included as predictors of child behavior problems. To obtain an estimate of the moderating role of child irritability and fearful distress prior to testing for the mediational role of child reactivity to conflict, all pathways involving mediational factors were constrained to 0 (i.e., not estimated) in the analyses. Correlations were also estimated among all predictors in the model, with the exception of child gender with family income. Only significant correlations are shown in the figure for the sake of clarity.

The model for the moderating role of children's temperamental irritability, which is depicted in Figure 3, provided an adequate representation of

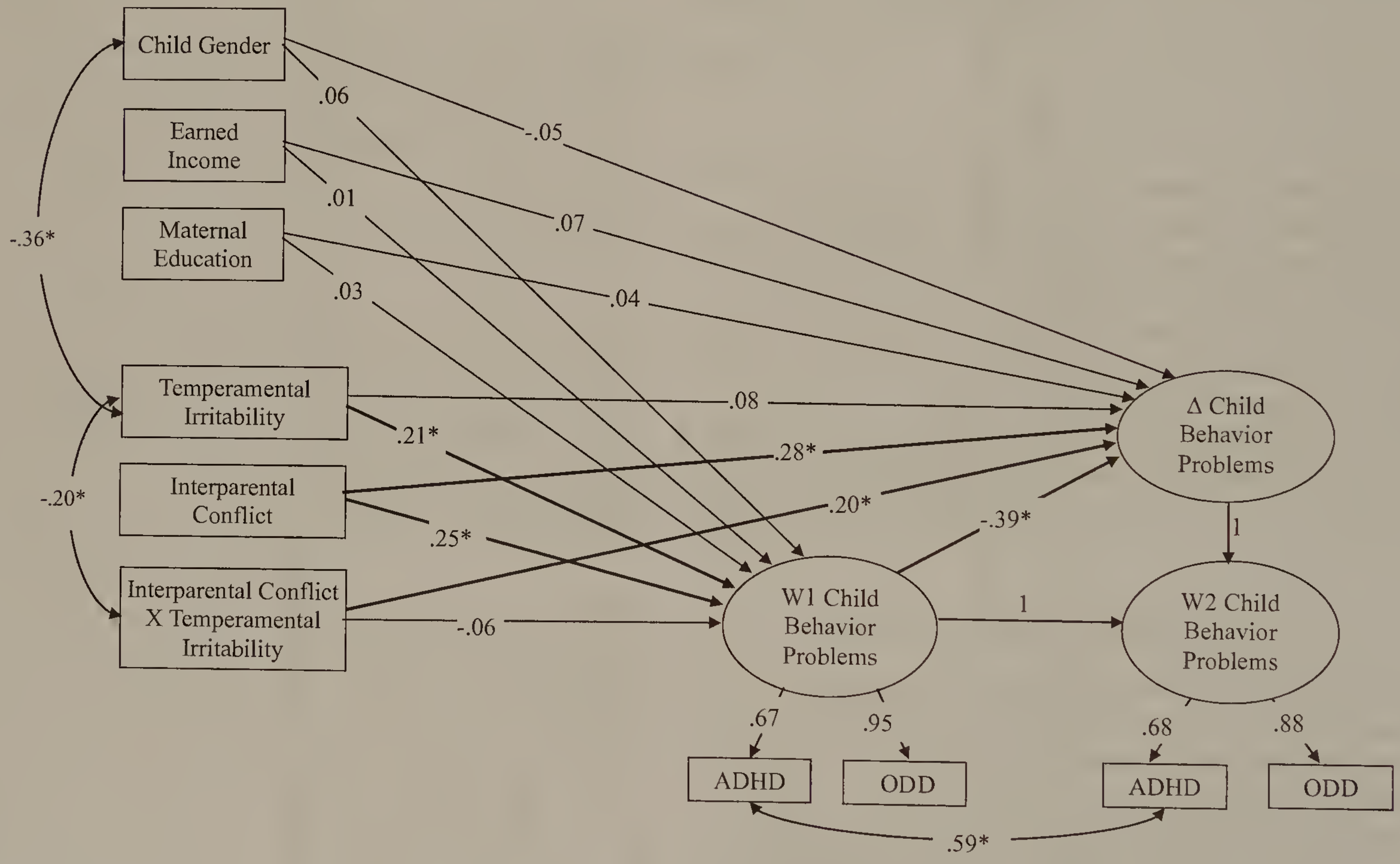

Child Gender
Earned Income
Maternal Education
Temperamental Irritability
Interparental Conflict
Interparental Conflict X Temperamental Irritability
Δ Child Behavior Problems
W1 Child Behavior Problems
W2 Child Behavior Problems
ADHD
ODD
ADHD
ODD
-.36*
-.20*
.06
-.05
.01
.07
.03
.04
.08
.21*
.28*
.20*
.25*
-.39*
-.06
1
1
.67
.95
.68
.88
.59*

Figure 3. Structural equation model examining the interaction between temperamental irritability and interparental conflict predicting child behavior problems. Bolded lines denote significant paths and nonbolded lines signify nonsignificant paths. All pathways involving the proposed mediational factors were constrained to 0 and are not shown in the model for sake of simplicity. Correlations were estimated among all predictors with the exception of child gender with earned income/social status. Only significant associations are shown for clarity. ADHD = attention deficit hyperactivity disorder; ODD = oppositional defiant disorder; W1 = Wave 1; W2 = Wave 2; Δ = change.
$^*p \leq .05$.

the data, $\chi^2(84, N = 201) = 172.81$, $p < .001$, root mean square error of approximation (RMSEA) = 0.07, χ^2/df ratio = 2.06, comparative fit index (CFI) = 0.92. Concurrent child behavior problems at Wave 1 were significantly predicted by irritability, $\beta = .21$, $p < .01$, and interparental conflict, $\beta = .25$, $p < .001$. Interparental conflict was also a significant predictor of change in disruptive behavior problems, $\beta = .28$, $p < .01$. Of relevance to our aims, the Interparental Conflict × Temperamental Irritability interaction predicted latent change in children's behavior problems over the two waves even after inclusion of the covariates as predictors, $\beta = .20$, $p < .05$.

To better characterize the interaction, we calculated the simple slopes of child behavior problems for children high and low (±1 *SD*) in temperamental irritability. The findings indicated that interparental conflict was significantly associated with latent change in child behavior problems for children high in irritability, $b = 1.36$, $p < .001$, but not for children low in irritability, $b = .12$, $p = .76$. Consistent with recommendations for adequately testing differential susceptibility (Roisman et al., 2012), the graphical plot of the interaction was depicted at 2 *SD* above and below the mean for interparental conflict. The resulting plot in Figure 4 revealed a disordinal interaction characterized by a cross-over of the two regression lines. In accord with differential susceptibility predictions, children higher in irritability fared worse in terms of behavior problems in destructive interparental contexts and better in supportive interparental contexts.

Due to difficulties in authoritatively interpreting the correspondence between graphical plots and moderator predictions derived from diathesis-stress and differential susceptibility models, Roisman et al. (2012) recommended the use of definitive quantitative indices to supplement conventional ways of dissecting interactions. Accordingly, we utilized the proportion affected (PA) index to determine the proportion of the sample of children who fall within the "for better" region of the graphical plot. Because the PA index is not dependent on sample size or variations in the specified range of the environmental predictor (e.g., 1 or 2 *SD*), it is regarded as a particularly robust quantitative analysis of the form of the interaction. Results of the calculations yielded a PA index of .37, a value well above the .16 threshold for providing greater support for differential susceptibility theory over the diathesis-stress model (Roisman et al., 2012). The PA index specifically indicates that 37% of the children fell below the crossover point of the interaction and within the "for better" region in which higher levels of child irritability are associated with greater decreases in behavior problems.

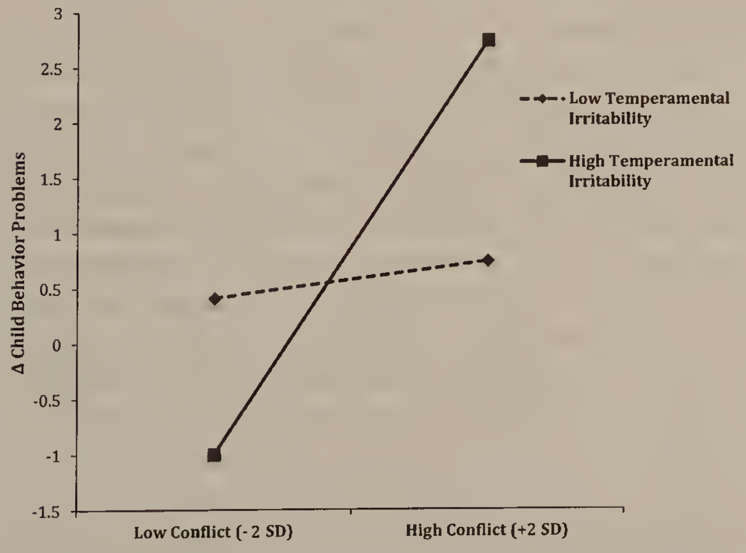


Figure 4. Simple slope plot of interaction between temperamental irritability and interparental conflict predicting latent change in child behavior problems, at ±2 *SD* of interparental conflict. Δ = latent change.

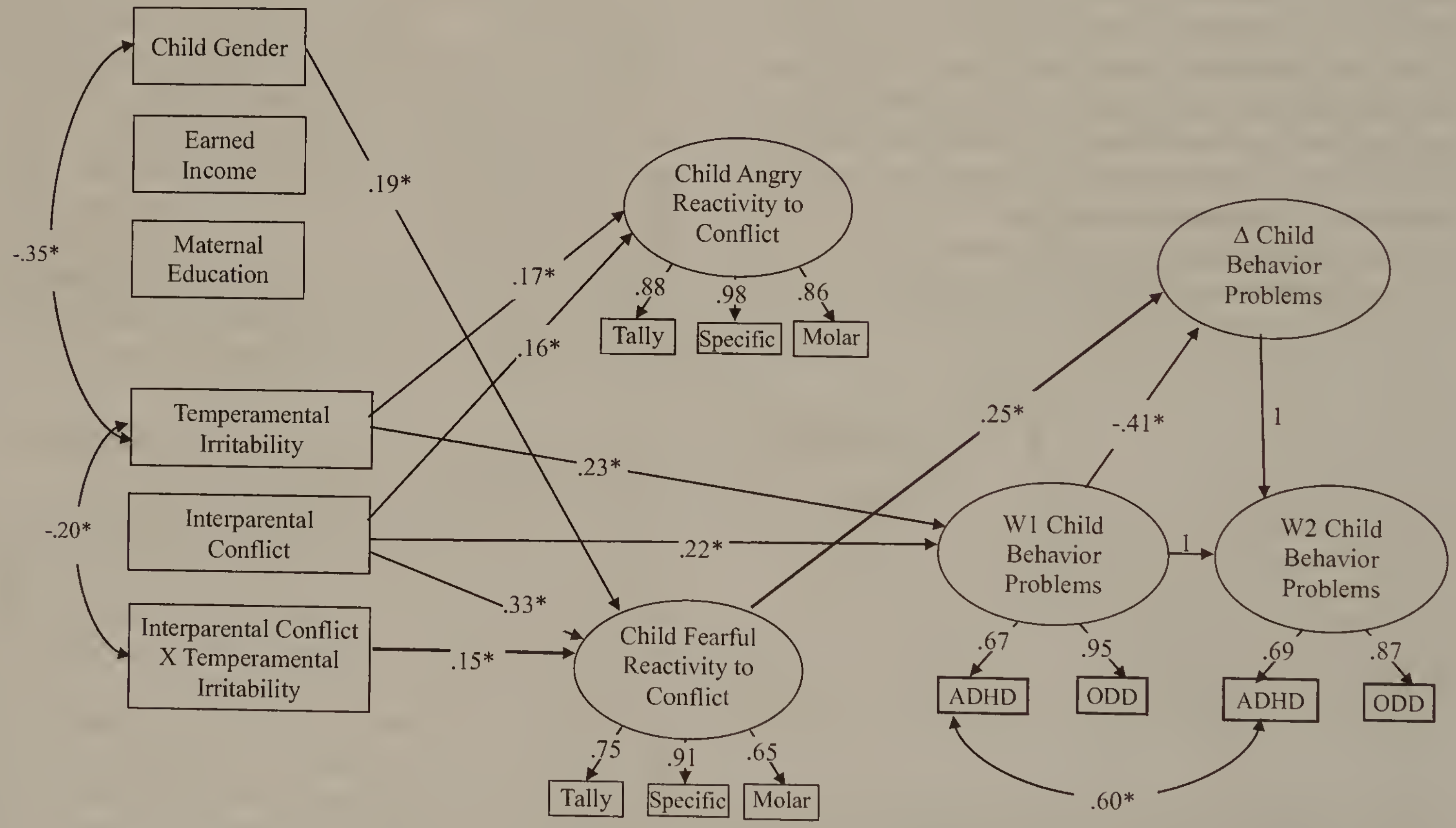
The model examining children's temperamental fearful distress as a moderator of the association between interparental conflict and child behavior problems also provided a satisfactory representation of the data, $\chi^2(84, N = 201) = 165.05$, $p < .001$, RMSEA = 0.07, χ^2/df ratio = 1.97, CFI = 0.92. Consistent with the temperamental irritability findings, interparental conflict significantly predicted concurrent behavior problems, $\beta = .25$, $p = .001$, and increases in behavior problems over the 1-year assessment period, $\beta = .27$, $p < .01$. Temperamental fearful distress, however, was not significantly associated with concurrent or prospective disruptive behavior problems in the model as either a predictor or a moderator of interparental conflict.

Child Emotional Responses to Conflict as a Mediator of the Moderating Effects of Irritability

Given the significant moderator results for irritability, the next objective was to examine why prospective associations between interparental conflict and child behavior problems varied significantly as a function of temperament. Therefore, we specified a mediated moderation model to test whether children's emotional responses to interparental conflict mediated the moderating role of temperamental irritability in the pathway between interparental conflict and child behavior problems. Support for a mediated moderation model requires demonstrating that: (a) the multiplicative product of interparental conflict and irritability significantly predicts children's emotional responses to conflict; (b) negative emotional responses to conflict, in turn, predict higher levels of subsequent child behavior problems; and (c) the resulting indirect pathway is statistically significant (Muller, Judd, & Yzerbyt, 2005). Figure 5 depicts the structural paths estimated to test the mediated moderation model with the inclusion of two forms of emotional responses to interparental conflict as the mediators: fearful and angry responses. In addition to specifying all the paths among the predictors and covariates and the outcomes from the first analytic step, we specifically estimated pathways running from all predictors (i.e., interparental conflict, irritability, their interaction term, and the covariates) to children's fearful and angry responding to interparental conflict. Angry and fearful responses to conflict, in turn, were specified as predictors of child behavior problems.

The resulting model provided a good representation of the data, $\chi^2(68, N = 201) = 117.11$, $p > .001$, RMSEA = 0.06, χ^2/df ratio = 1.72, CFI = 0.96. Angry responses to interparental conflict did not meet criteria for mediation due to its negligible links with the interparental Conflict × Irritability Interaction term, $\beta = -.06$, *ns*, and change in child behavior problems, $\beta = .10$, *ns*. However, the interaction between irritability and interparental conflict significantly predicted fearful reactivity, $\beta = .15$, $p < .05$. Fearful reactivity, in turn, predicted latent change in child behavior problems, $\beta = .25$, $p < .05$. In further support of mediated moderation of fearful reactivity, bootstrapping tests using the PRODCLIN software program indicated the indirect effect path involving the Temperamental Irritability × Interparental Conflict interaction, fearful reactivity to conflict, and latent change in children's behavior problems was significant, 95% CI [0.001,0.339] (MacKinnon, Fritz, Williams, & Lockwood, 2007).

To portray the form of the moderation effects of irritability in the association between interparental conflict and children's fearful reactivity to interparental conflict, we followed the same procedures used in characterizing the significant interaction predicting child behavior problems. Simple slope analyses indicated that interparental conflict significantly predicted fearful reactivity for children who were high (+1 *SD*) in temperamental irritability, $b = .82$, $p < .001$, but not low in irritability, $b = .27$, $p = .14$. Figure 6 depicts the graphical plot within



Child Gender
Earned Income
Maternal Education
Temperamental Irritability
Interparental Conflict
Interparental Conflict X Temperamental Irritability
-.35*
-.20*
.19*
.17*
.16*
.23*
.22*
.33*
.15*
Child Angry Reactivity to Conflict
.88
.98
.86
Tally
Specific
Molar
Child Fearful Reactivity to Conflict
.75
.91
.65
Tally
Specific
Molar
Δ Child Behavior Problems
.25*
-.41*
1
W1 Child Behavior Problems
1
W2 Child Behavior Problems
.67
.95
ADHD
ODD
.69
.87
ADHD
ODD
.60*

Figure 5. Structural equation model examining whether the moderating role of temperamental irritability in the link between interparental conflict and change in child behavior problems is mediated by child reactivity to conflict. Pathways were estimated between all predictors, but only significant pathways are shown for clarity. ADHD = attention deficit hyperactivity disorder; ODD = oppositional defiant disorder; W = Wave; Δ = change.
*$p \le .05$.

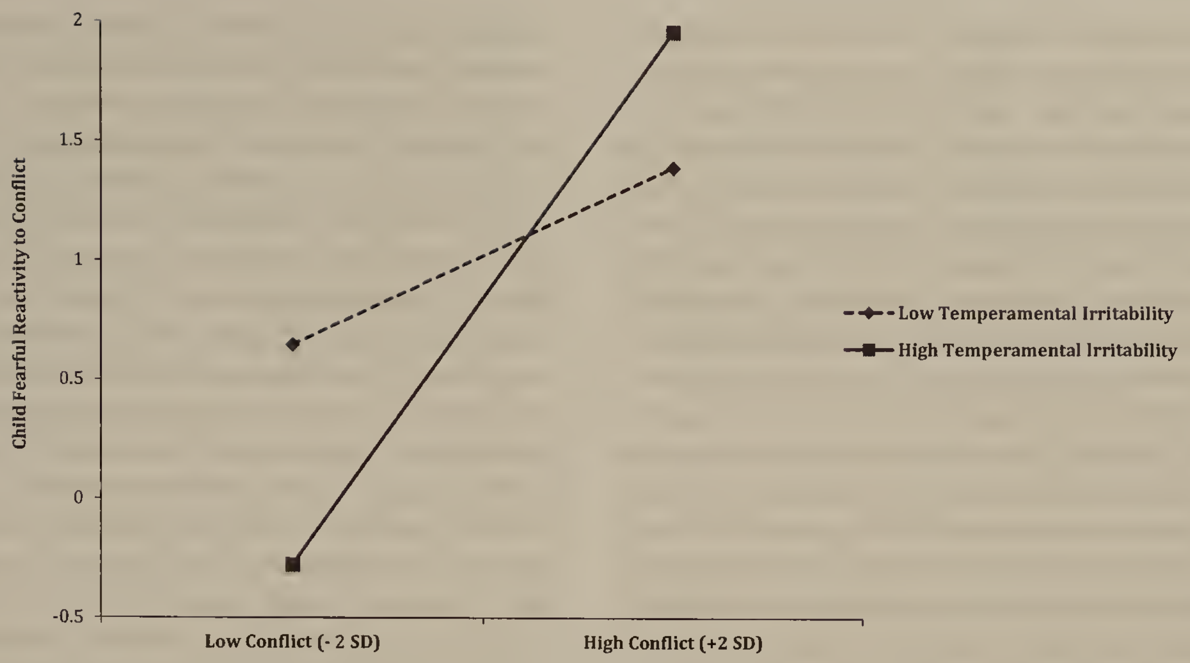


Figure 6. Simple slope plot of interaction between temperamental irritability and interparental conflict predicting child fearful reactivity to conflict, at ±2 *SD* of interparental conflict.

the range of 2 *SD* around the mean of interparental conflict. Visual inspection of the disordinal plot provides some initial support for differential susceptibility theory. Consistent with the plot, the PA index, which assesses the proportion of children in the "for better" range of the plot, was .72. High irritability was specifically associated with lower fearful reactivity for 72% of the range of experiences with interparental conflict. Thus, the PA index offers stronger support for differential susceptibility over the diathesis-stress model.

Discussion

Our study was designed to extend earlier findings indicating that interparental conflict only predicted greater behavior problems for infants who were higher in negative emotionality in several ways (Pauli-Pott & Beckmann, 2007). First, to address the increasing differentiation in forms of negative emotionality that children experience during the early preschool years, a main objective was to disentangle temperamental irritability and fearful distress as moderators of prospective associations between interparental conflict and preschool disruptive behavior problems over a 1-year period. Second, our study was designed to specifically test the relative correspondence of the significant moderating effects of temperament with predictions of diathesis-stress and differential susceptibility theories (Belsky, 2005). Third, to understand why the risk posed by interparental conflict to children varies as a function of their level of negative emotionality, our final aim was to test whether children's emotional responses to interparental conflict mediated the moderating effects of negative emotionality. The results indicated that temperamental irritability moderated the prospective path between interparental conflict and child behavior problems in a manner supported by differential susceptibility theory (Belsky & Pluess, 2009). Mediated moderation analyses further revealed that fearful responses to interparental conflict partially accounted for why children with high temperamental irritability evidenced greater plasticity across supportive and adverse interparental contexts.

Increasing calls to distinguish between forms of negative emotionality have largely been directed toward dissecting temperamental irritability and fearful distress as differential risk factors for forms of psychopathology (Rothbart, 2011; Sanson, Hemphill, & Smart, 2004). In support of increasing specificity in the conceptualization of negative emotionality, our results showed that the association between a multimethod composite of interparental conflict and increases in child disruptive behavior problems over a 1-year period was moderated specifically by children's temperamental irritability rather than their fearful distress. Follow-up analyses designed to characterize the form of moderation provided strong support for differential susceptibility theory and its designation of temperamental irritability as a plasticity factor rather than a diathesis. Children high in temperamental irritability specifically experienced outcomes in a "for better" and "for worse" manner hypothesized by differential susceptibility theory. In the "for worse" part of the interaction, children with high temperamental irritability exhibited greater increases in behavior problems over a 1-year period when exposed to high destructive interparental conflict. Conversely, in the "for better" part of the equation, temperamentally irritable children evidenced greater reductions in behavior problems relative to children low in irritability under conditions of constructive interparental conflict.

These findings beg the question of why temperamental irritability serves as a plasticity factor in the context of interparental conflict. Given that conceptualizations of interparental conflict and temperament each underscores the significance of children's emotional responding to stressful events as pivotal for understanding individual differences in their psychological adjustment (Davies & Cummings, 1994; Zalewski, Lengua, Wilson, Trancik, & Bazinet, 2011), we examined whether children's fearful and angry responses to interparental conflict mediated the moderating role of temperamental irritability in the pathway between interparental conflict and children's behavior problems. Although interparental conflict and temperamental irritability each predicted children's angry responding to interparental conflict, mediated moderation was not supported due to nonsignificant paths between (a) the interaction involving interparental conflict and temperamental irritability and children's angry reactivity, and (b) children's angry reactivity and their subsequent behavior problems. In contrast, tests of mediated moderation supported children's fearful reactivity to interparental conflict as a mediator of the moderating role of temperamental irritability. The Interparental Conflict × Temperamental Irritability interaction was specifically associated with children's fearful reactivity to conflict, which in turn predicted subsequent increases in their disruptive behavior problems 1 year later. Although caution should be exercised in interpreting these findings until they are replicated, these findings support emotional security theory and its hypothesis that fearful reactivity to family adversity is a primary mechanism mediating the interplay of interparental discord and child temperament and their psychological adjustment (e.g., Davies & Sturge-Apple, 2007).

In the first link of the mediated-moderation chain, follow-up tests of the moderating role of temperamental irritability decisively favored predictions from differential susceptibility theory over the diathesis-stress model. On the "for worse" side of the spectrum, irritable children evidenced higher fearful reactivity to conflict when exposed to heightened destructive interparental conflict. However, in the "for better" part of the spectrum, higher irritability was associated with significantly lower fearful reactivity to conflict in contexts of constructive interparental conflict. As one explanation, it is possible that temperamental irritability is an index of greater sensitivity to environmental cues. For example, higher sensory sensitivity to environmental stimuli has been conceptualized as a primary mechanism underlying individual differences in plasticity (e.g., Belsky & Pluess, 2009).

However, the noticeably larger "for better" portion of the interaction (i.e., 72% PA index) is not fully consistent with the notion that children with temperamental irritability exhibit comparable sensitivity to adverse and supportive interparental stimuli. As a complementary interpretation, other evolutionary theories of temperament postulate that temperamental irritability may have persisted over evolutionary time by increasing bold, aggressive, and assertive strategies for accessing resources during bouts of interpersonal competition, challenge, or conflict (Korte et al., 2005; Réale, Reader, Sol, McDougall, & Dingemanse, 2007). The success of this strategy in challenging social contexts depends on inhibiting the experience and expression of signs of vulnerability, such as fear and distress, in favor of more dominant emotions, like anger and aggression (Davies & Sturge-Apple, 2007). Thus, children with irritable temperaments may experience dampened fear when bouts of interparental conflict range from benign to moderately threatening. However, in highly threatening contexts where risk of injury is high, evolutionary theories posit that bold, irritable individuals evidence demonstrative displays of fearful reactivity with the enactment of "flight" patterns of response (Gilbert, 2001; Plutchik, 2001). Consistent with this interpretation, the current study found that irritable children only showed greater susceptibility to fearful responses within a narrow range of the highest levels of interparental conflict.

In the second link in the mediated moderation chain, fearful reactivity to interparental conflict predicted subsequent increases in children's behavior problems. These findings are consistent with conceptualizations that highlight fearful reactivity and appraisals of threat as unique predictors of children's psychological problems (Crockenberg & Langrock, 2001; Davies & Cummings, 1994; Grych & Fincham, 1990). For example, emotional security theory proposes that prolonged fearful reactivity reflects protracted concerns with emotional security that increase the prioritization of defending against interpersonal threat at the expense of other significant developmental goals. Fear and insecurity in interparental contexts have been shown to increase children's vulnerability to psychological problems by undermining attention and effortful control processes (see review by Davies, Sturge-Apple, & Martin, 2013). Resulting depletions in self-regulatory abilities and attention, in turn, are further postulated to increase children's disruptive problems (e.g., Davies & Sturge-Apple, 2007). Consistent with this hypothesis, research has shown that children's fearful reactivity to interparental conflict is a unique predictor of children's behavior problems even when taking into account other forms of reactivity (Davies et al., 2012).

Differences in the sizes of the "for better" regions of temperamental irritability for children's fearful reactivity to conflict and changes in their behavior problems also have important implications in interpreting our findings. According to the PA index, the proportion of the sample in which irritability is associated with improved outcomes is noticeably higher in the prediction of fearful responses to interparental conflict (i.e., .72) than in the prediction of disruptive behavior problems (i.e., .37). These findings suggest that the mediating role of fearful reactivity is more valuable in explaining why children high in temperamental irritability evidence decreases in behavior problems when interparental conflict is low. Thus, other mechanisms beyond children's fearful reactivity to conflict may also be operating to give rise to differential susceptibility in children's disruptive behavior problems. Identifying the processes underlying the "for worse" part of temperamental irritability is an important direction for future research.

Several limitations in our study also merit discussion. First, interparental conflict, children's responses to interparental conflict, and their behavior problems were all derived in some form from maternal reports. Our multimethod approach and use of latent analyses for indexing change in children's behavior problems reduces some concern about the operation of common-method bias, especially in the context of our primary focus on testing the moderator effects of temperament. However, to further address this limitation, design improvements in future research might be achieved through more intensive multimethod, multi-informant approaches. Second, although we employed a longitudinal design to capture changes in child behavior problems, using two measurement occasions does not permit a full, prospective analysis of change at each link in the mediated moderation model (Cole & Maxwell, 2003). Third, because our tests of mediated moderation were conducted with a sample of families who experienced elevated interparental aggression and socioeconomic disadvantage, caution should be exercised in generalizing these findings to other (e.g., more advantaged) samples.

In addition, our measurement of temperamental irritability was derived from procedures that were designed to assess children's reactions to novelty, risk, and unexpected events. Thus, replicating our findings using a more inclusive measurement battery that captures children's reactions to a broader set of challenges (e.g., blocked goals) is an important direction for future research. Procedures that do not naturally evoke high levels of frustration may more sensitively capture a specific dimension of temperamental irritability (e.g., contextually inappropriate anger). For example, in the temperamental fear literature, expressions of fear in relatively low threat tasks are stronger predictors of subsequent anxiety problems than fear displays in high threat contexts (Buss, 2011). By the same token, evidence supports the validity of our assessment of temperamental irritability measure. For example, observations of anger and frustration across a wide array of novel and challenging contexts have been shown to cohere together to form an internally consistent composite of temperamental irritability. In further supporting the validity of our measurement approach, indices of children's disruptive behavior problems were consistently associated with temperamental irritability but not fearful distress.

Finally, although we did not find evidence for the moderating role of temperamental fearful distress, more research is needed before we can draw any definitive conclusions about its role in altering associations between interparental conflict and child adjustment. An unanswered question within differential susceptibility theory is whether plasticity markers (e.g., forms of negative emotionality) are domain general or if they may differ in regard to specific aspects of functioning (Belsky, 2005; Belsky & Pluess, 2009). Children high in fearful distress are posited to be at greater risk for disorders of overcontrol (Rothbart, 2011); thus, a derivative hypothesis from a more domain specific perspective is that temperamental fearful distress might moderate interparental conflict in models of other forms of psychological adjustment, such as internalizing symptoms. In addition, complementary advances in understanding why fearful reactivity mediates the moderating effects of temperamental irritability may be achieved through systematically characterizing the complexity of associations between temperamental irritability and children's fearful reactivity to interparental conflict.

In conclusion, our study was designed to break new ground in several ways. As a first foray into differentiating between forms of negative emotionality as moderators of interparental conflict, our findings indicate that prospective associations between interparental conflict and children's behavior problems specifically varied as a function of temperamental irritability. Follow-up tests further indicated that the form of moderation decidedly favored hypotheses derived from differential susceptibility theory over the diathesis-stress model. Toward further advancing an understanding of why temperamental irritability conferred significant advantages in supportive interparental conflicts and substantial costs in highly adverse developmental contexts, additional multivariate analyses revealed that fearful reactivity to conflict mediated the moderating effects of temperamental irritability. Moreover, the mediating role of fearful reactivity appears to provide a better explanation for why children high in temperamental irritability experience improved outcomes in contexts of interparental harmony than why they exhibit substantial behavior problems in the face of destructive interparental relationships.

References

Achenbach, T. M., & Rescorla, L. A. (2006). The Achenbach system of empirically based assessment. In R. P. Archer (Ed.), *Forensic uses of clinical assessment instruments* (pp. 229–262). Mahwah, NJ: Erlbaum.

Aebi, M., Metzke, C. W., & Steinhausen, H. C. (2010). Accuracy of the DSM-oriented attention problem scale of the child behavior checklist in diagnosing attention-deficit hyperactivity disorder. *Journal of Attention Disorders, 13*, 454–463. doi:10.1177/1087054708325739

Arbuckle, J. L. (2009). *Amos 18.0*. Chicago, IL: SPSS..

Bellina, M., Brambilla, P., Garzitto, M., Negri, G. A., Molteni, M., & Nobile, M. (2013). The ability of CBCL DSM-oriented scales to predict DSM–IV diagnoses in a referred sample of children and adolescents. *European Child & Adolescent Psychiatry, 22*, 235–246. doi:10.1007/s00787-012-0343-0

Belsky, J. (2005). Differential susceptibility to rearing influence: An evolutionary hypothesis and some evidence. In B. J. Ellis & D. F. Bjorklund (Eds.), *Origins of the social mind: Evolutionary psychology and child development* (pp. 139–163). New York, NY: Guilford.

Belsky, J., & Pluess, M. (2009). Beyond diathesis stress: Differential susceptibility to environmental influences. *Psychological Bulletin, 135*, 885–908. doi:10.1037/a0017376

Blair, K. A., Denham, S. A., Kochanoff, A., & Whipple, B. (2004). Playing it cool: Temperament, emotion regulation, and social behavior in preschoolers. *Journal of School Psychology, 42*, 419–443. doi:10.1016/j.jsp.2004.10.002

Braungart-Rieker, J. M., Hill-Soderlund, A. L., & Karrass, J. (2010). Fear and anger reactivity trajectories from 4 to 16 months: The roles of temperament, regulation, and maternal sensitivity. *Developmental Psychology, 46*, 791–804. doi:10.1037/a0019673

Buss, K. A. (2011). Which fearful toddlers should we worry about? Context, fear regulation, and anxiety risk. *Developmental Psychology, 47*, 804–819. doi:10.1037/a0023227

Buss, K. A., & Goldsmith, H. H. (1998). Fear and anger regulation in infancy: Effects on the temporal dynamics of affective expression. *Child Development, 69*, 359–374. doi:10.1111/j.1467-8624.1998.tb06195.x

Calkins, S. D., Dedmon, S. E., Gill, K. L., Lomax, L. E., & Johnson, L. M. (2002). Frustration in infancy: Implications for emotion regulation, physiological processes, and temperament. *Infancy, 3*, 175–197. doi:10.1207/S15327078IN0302_4

Colder, C. R., Lochman, J. E., & Wells, K. C. (1997). The moderating effects of children's fear and activity level on relations between parenting practices and childhood symptomatology. *Journal of Abnormal Child Psychology, 25*, 251–263. doi:10.1023/A:1025704217619

Cole, D. A., & Maxwell, S. E. (2003). Testing mediational models with longitudinal data: Questions and tips in the use of structural equation modeling. *Journal of Abnormal Psychology, 112*, 558–577. doi:10.1037/0021-843X.112.4.558

Crockenberg, S., & Langrock, A. (2001). The role of emotion and emotional regulation in children's responses to interparental conflict. In J. H. Grych & F. D. Fincham (Eds.), *Interparental conflict and child development: Theory, research, and applications* (pp. 129–156). New York, NY: Cambridge University Press.

Cummings, E. M., & Davies, P. T. (2010). *Marital conflict and children: An emotional security perspective*. New York, NY: Guilford.

Cunningham, C. E., & Boyle, M. H. (2002). Preschoolers at risk for attention-deficit hyperactivity disorder and oppositional defiant disorder: Family, parenting, and behavioral correlates. *Journal of Abnormal Child Psychology, 30*, 555–569. doi:10.1023/A:1020855429085

Davies, P. T., Cicchetti, D., & Martin, M. J. (2012). Toward greater specificity in identifying associations among interparental aggression, child emotional reactivity to conflict, and child problems. *Child Development, 83*, 1789–1804. doi:10.1111/j.1467-8624.2012.01804.x

Davies, P. T., & Cummings, E. M. (1994). Marital conflict and child adjustment: An emotional security hypothesis. *Psychological Bulletin, 116,* 387–411. doi:10.1037/0033-2909.116.3.387

Davies, P. T., & Sturge-Apple, M. L. (2007). Advances in the formulation of emotional security theory: An ethologically based perspective. *Advances in Child Development and Behavior, 35,* 87–137. doi:10.1016/B978-0-12-009735-7.50008-6

Davies, P. T., Sturge-Apple, M. L., & Martin, M. J. (2013). Family discord and child health: An emotional security formulation. In N. S. Landale, S. M. McHale, & A. Booth (Eds.), *Families and child health* (pp. 45–74). New York, NY: Springer.

Davies, P., & Windle, M. (2001). Interparental discord and adolescent adjustment trajectories: The potentiating and protective role of intrapersonal attributes. *Child Development, 72,* 1163–1178. doi:10.1111/1467-8624.00340

Derryberry, D., & Rothbart, M. K. (1997). Reactive and effortful processes in the organization of temperament. *Development and Psychopathology, 9,* 633–652.

Durbin, C. E., Klein, D. N., Hayden, E. P., Buckley, M. E., & Moerk, K. C. (2005). Temperamental emotionality in preschoolers and parental mood disorders. *Journal of Abnormal Psychology, 114,* 28–37. doi:10.1037/0021-843X.114.1.28

Emery, R. E. (1989). Family violence. *American Psychologist, 44,* 321–328. doi:10.1037/0003-066X.44.2.321

Fantuzzo, J., Boruch, R., Beriama, A., Atkins, M., & Marcus, S. (1997). Domestic violence and children: Prevalence and risk in five major U.S. cities. *Journal of the American Academy of Child & Adolescent Psychiatry, 36,* 116–122. doi:10.1097/00004583-199701000-00025

Fergusson, D. M., & Horwood, L. J. (1998). Exposure to interparental violence in childhood and psychosocial adjustment in young adulthood. *Child Abuse & Neglect, 22,* 339–357. doi:10.1016/S0145-2134(98)00004-0

Fincham, F. D. (1994). Understanding the association between marital conflict and child adjustment: Overview. *Journal of Family Psychology, 8,* 123–127. doi:10.1037/0893-3200.8.2.123

Fox, N. A., Henderson, H. A., Rubin, K. H., Calkins, S. D., & Schmidt, L. A. (2001). Continuity and discontinuity of behavioral inhibition and exuberance: Psychophysiological and behavioral influences across the first four years of life. *Child Development, 72,* 1–21. doi:10.1111/1467-8624.00262

Gilbert, P. (2001). Evolutionary approaches to psychopathology: The role of natural defences. *Australian and New Zealand Journal of Psychiatry, 35,* 17–27. doi:10.1046/j.1440-1614.2001.00856.x

Goldstein, L. H., Harvey, E. A., Friedman-Weieneth, J. L., Pierce, C., Tellert, A., & Sippel, J. C. (2007). Examining subtypes of behavior problems among 3-year-old children, part II: Investigating differences in parent psychopathology, couple conflict, and other family stressors. *Journal of Abnormal Child Psychology, 35,* 111–123. doi:10.1007/s10802-006-9088-x

Grych, J. H., & Fincham, F. D. (1990). Marital conflict and children's adjustment: A cognitive-contextual framework. *Psychological Bulletin, 108,* 267–290. doi:10.1037/0033-2909.108.2.267

Hayden, E. P., Klein, D. N., Durbin, C. E., & Olino, T. M. (2006). Positive emotionality at age 3 predicts cognitive styles in 7-year-old children. *Development and Psychopathology, 18,* 409–423. doi:10.1017/S0954579406060226

Ingram, R. E., & Luxton, D. D. (2005). Vulnerability-stress models. In B. L. Hankin & J. R. Abela (Eds.), *Development of psychopathology: A vulnerability-stress perspective* (pp. 32–46). Thousand Oaks, CA: Sage.

Kagan, J., Reznick, J. S., & Snidman, N. (1987). The physiology and psychology of behavioral inhibition in children. *Child Development, 58,* 1459–1473.

Kerig, P. K. (1996). Assessing the links between interparental conflict and child adjustment: The conflicts and problem-solving scales. *Journal of Family Psychology, 10,* 454–473. doi:10.1037/0893-3200.10.4.454

Kitzmann, K. M., Gaylord, N. K., Holt, A. R., & Kenny, E. D. (2003). Child witnesses to domestic violence: A meta-analytic review. *Journal of Consulting and Clinical Psychology, 71,* 339–352. doi:10.1037/0022-006X.71.2.339

Korte, S. M., Koolhaas, J. M., Wingfield, J. C., & McEwen, B. S. (2005). The Darwinian concept of stress: Benefits of allostasis and costs of allostatic load and the trade-offs in health and disease. *Neuroscience and Biobehavioral Reviews, 29,* 3–38. doi:10.1016/j.neubiorev.2004.08.009

LeDoux, J. E. (2000). Emotion circuits in the brain. *Annual Review of Neuroscience, 23,* 155–184. doi:10.1146/annurev.neuro.23.1.155

Lemery, K. S., Goldsmith, H. H., Klinnert, M. D., & Mrazek, D. A. (1999). Developmental models of infant and childhood temperament. *Developmental Psychology, 35,* 189–204. doi:10.1037/0012-1649.35.1.189

Lengua, L. J. (2008). Anxiousness, frustration, and effortful control as moderators of the relation between parenting and adjustment in middle-childhood. *Social Development, 17,* 554–577. doi:10.1111/j.1467-9507.2007.00438.x

MacKinnon, D. P., Fritz, M. S., Williams, J., & Lockwood, C. M. (2007). Distribution of the product confidence limits for the indirect effect: Program PRODCLIN. *Behavior Research Methods, 39,* 384–389. doi:10.3758/BF03193007

McArdle, J. J. (2009). Latent variable modeling of differences and changes with longitudinal data. *Annual Review of Psychology, 60,* 577–605. doi:10.1146/annurev.psych.60.110707.163612

McLoyd, V. C., Harper, C. I., & Copeland, N. L. (2001). Ethnic minority status, interparental conflict, and child adjustment. In J. H. Grych & F. D. Fincham (Eds.), *Interparental conflict and child development: Theory, research, and application* (pp. 98–125). New York, NY: Cambridge University Press.

Morris, A. S., Silk, J. S., Steinberg, L., Sessa, F. M., Avenevoli, S., & Essex, M. J. (2002). Temperamental vulnerability and negative parenting as interacting predictors of child adjustment. *Journal of Marriage and Family, 64*, 461–471. doi:10.1111/j.1741-3737.2002.00461.x

Muller, D., Judd, C. M., & Yzerbyt, V. Y. (2005). When moderation is mediated and mediation is moderated. *Journal of Personality and Social Psychology, 89*, 852–863. doi:10.1037/0022-3514.89.6.852

Nakamura, B. J., Ebesutani, C., Bernstein, A., & Chorpita, B. F. (2009). A psychometric analysis of the child behavior checklist DSM-oriented scales. *Journal of Psychopathology and Behavioral Assessment, 31*, 178–189. doi:10.1007/s10862-008-9119-8

Pauli-Pott, U., & Beckmann, D. (2007). On the association of interparental conflict with developing behavioral inhibition and behavior problems in early childhood. *Journal of Family Psychology, 21*, 529–532. doi:10.1037/0893-3200.21.3.529

Plutchik, R. (2001). The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American Scientist, 89*, 344–350.

Poehlmann, J., Schwichtenberg, A. J. M., Shlafer, R. J., Hahn, E., Bianchi, J., & Warner, R. (2011). Emerging self-regulation in toddlers born preterm or low birth weight: Differential susceptibility to parenting? *Development and Psychopathology, 23*, 177. doi:10.1017/S0954579410000726

Réale, D., Reader, S. M., Sol, D., McDougall, P. T., & Dingemanse, N. J. (2007). Integrating animal temperament within ecology and evolution. *Biological Reviews, 82*, 291–318. doi:10.1111/j.1469-185X.2007.00010.x

Roisman, G. I., Newman, D. A., Fraley, R. C., Haltigan, J. D., Groh, A. M., & Haydon, K. C. (2012). Distinguishing differential susceptibility from diathesis–stress: Recommendations for evaluating interaction effects. *Development and Psychopathology, 24*, 389–409. doi:10.1017/S0954579412000065

Rothbart, M. K. (2011). *Becoming who we are: Temperament and personality in development*. New York, NY: Guilford.

Sanson, A., Hemphill, S. A., & Smart, D. (2004). Connections between temperament and social development: A review. *Social Development, 13*, 142–170. doi:10.1046/j.1467-9507.2004.00261.x

Shaw, D. S., Bell, R. Q., & Gilliom, M. (2000). A truly early starter model of antisocial behavior revisited. *Clinical Child and Family Psychology Review, 3*, 155–172. doi:10.1023/A:1009599208790

Spjeldnes, S., & Choi, J. K. (2008). Direct and indirect effects of interparental relationship quality on child behavior problems in low-income, Black, single-mother families. *Marriage & Family Review, 44*, 411–438. doi:10.1080/01494920802453910

Straus, M. A., Hamby, S. L., Boney-McCoy, S., & Sugarman, D. B. (1996). The Revised Conflict Tactics Scales (CTS2): Development and preliminary psychometric data. *Journal of Family Issues, 17*, 283–316. doi:10.1177/019251396017003001

Van der Mark, I. L., Bakermans-Kranenburg, M. J., & van IJzendoorn, M. H. (2002). The role of parenting, attachment, and temperamental fearfulness in the prediction of compliance in toddler girls. *British Journal of Developmental Psychology, 20*, 361–378. doi:10.1348/026151002320620299

Waters, S. F., & Thompson, R. A. (2014). Children's perceptions of the effectiveness of strategies for regulating anger and sadness. *International Journal of Behavioral Development, 38*, 174–181. doi:10.1177/0165025413515410

Zalewski, M., Lengua, L. J., Wilson, A. C., Trancik, A., & Bazinet, A. (2011). Emotion regulation profiles, temperament, and adjustment problems in preadolescents. *Child Development, 82*, 951–966. doi:10.1111/j.1467-8624.2011.01575.x

Child Development, September/October 2015, Volume 86, Number 5, Pages 1351–1370

24-Month-Old Children With Larger Oral Vocabularies Display Greater Academic and Behavioral Functioning at Kindergarten Entry

Paul L. Morgan
The Pennsylvania State University

George Farkas
University of California, Irvine

Marianne M. Hillemeier
The Pennsylvania State University

Carol Scheffner Hammer
Teachers College, Columbia University

Steve Maczuga
The Pennsylvania State University

Data were analyzed from a population-based, longitudinal sample of 8,650 U.S. children to (a) identify factors associated with or predictive of oral vocabulary size at 24 months of age and (b) evaluate whether oral vocabulary size is uniquely predictive of academic and behavioral functioning at kindergarten entry. Children from higher socioeconomic status households, females, and those experiencing higher quality parenting had larger oral vocabularies. Children born with very low birth weight or from households where the mother had health problems had smaller oral vocabularies. Even after extensive covariate adjustment, 24-month-old children with larger oral vocabularies displayed greater reading and mathematics achievement, increased behavioral self-regulation, and fewer externalizing and internalizing problem behaviors at kindergarten entry.

Children with greater academic and behavioral functioning at kindergarten entry often experience better educational and societal opportunities as they age (Duncan et al., 2007; Sabol & Pianta, 2012). For example, children entering kindergarten with greater reading and mathematics achievement are more likely to attend college and enroll in higher quality institutions. They are also more likely to own homes, have 401(k) savings, be married, and live in higher income neighborhoods as adults (Chetty et al., 2011). Identifying factors contributing to greater academic and behavioral functioning at kindergarten entry should help guide efforts to deliver early interventions to specific population subgroups at risk for lower school functioning (Hoff, 2013; Lesaux, 2012), and so reduce later achievement gaps and increase postsecondary education, employment, productivity, and long-term wages (e.g., Heckman & Masterov, 2007).

Oral vocabulary is a malleable factor repeatedly theorized to contribute to increased academic and behavioral functioning, and so might be targeted in early interventions (e.g., Dickinson, Golinkoff, & Hirsh-Pasek, 2010; Perfetti & Stafura, 2014; see Appendix S1 in the online Supporting Information for further discussion of oral vocabulary terminology). Oral vocabulary refers to the words children use when speaking or recognize when listening. At-risk children's oral vocabularies have been theorized to constitute a "first-order" causal factor that, if increased, may enhance their developmental trajectories (Dickinson et al., 2010; Lesaux, 2012).

Oral Vocabulary's Theorized Relation With Academic and Behavioral Functioning

Theoretically, at about 24 months of age, accelerating growth in children's lexicons overtaxes their

---

Funding for this study was provided by the National Center for Special Education Research, Institute of Education Sciences, U.S. Department of Education (R324A120046). Infrastructure support was provided by the Pennsylvania State University's Population Research Institute through funding from the National Institute of Child Health and Human Development, National Institutes of Health (R24HD041025). No official endorsement should thereby be inferred. We thank Bruce Tomblin, Philip Dale, and Betsy Crais for their thoughtful suggestions on earlier versions of this manuscript.

Correspondence concerning this article should be addressed to Paul L. Morgan, Department of Education Policy Studies, 300 Rackley Building, the Pennsylvania State University, University Park, PA 16802. Electronic mail may be sent to paulmorgan@psu.edu.

© 2015 The Authors
Child Development © 2015 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2015/8605-0004
DOI: 10.1111/cdev.12398

protosyllabary (i.e., meaningless speech–motor patterns) and their use of word gestures (Levelt, Roelofs, & Meyer, 1999). To more effectively communicate, children are believed to rely increasingly on (a) phonemization, so that words begin to be represented by phonological segments, and (b) syntactization, in which lexical concepts begin to be grouped by categories and subcategories. Use of this dual-coding process, which is not necessarily contingent on the "vocabulary spurt" that itself may not occur for most children (Ganger & Brent, 2004) and can be explained by nonspecialized cognitive processes (McMurray, 2007), should allow young children to begin associating particular phonetic encodings and articulations with specific lexical representations, thereby facilitating acquisition and use of an oral vocabulary of increasingly greater size and complexity (Ouellette, 2006).

Acquiring a larger oral vocabulary should over time result in greater reading achievement (National Institute of Child Health and Human Development [NICHD] Early Child Care Research Network, 2005; Perfetti & Stafura, 2014; Walker, Greenwood, Hart, & Carta, 1994) by facilitating listening comprehension (Hoover & Gough, 1990) as well as decoding skills (Verhoeven, Leeuwe, & Vermeer, 2011). Children with larger oral vocabularies should more efficiently store words in their lexicons as smaller segments, helping to increase their phonological sensitivity and, thus, their understanding of the alphabetic principle and decoding (Silven, Poskiparta, Niemi, & Voeten, 2007). A larger oral vocabulary should also result in more efficient word identification (Perfetti & Hart, 2001), allowing children to better identify and understand partially decoded, irregular or novel, or orthographically complex words, and to infer spelling–sound relations (Mitchell & Brady, 2013). (Eventually, the relation between oral vocabulary and decoding becomes bidirectional, as independent reading results in new vocabulary acquisition.)

A larger oral vocabulary is also theorized to result in greater mathematics achievement (LeFevre et al., 2010). Children's initial understanding of the symbolic number system and its interrelations (e.g., number sequence, basic arithmetic and subtraction) is thought to rely on language-based verbal associations (e.g., Spelke & Tsivkin, 2001). Having a larger oral vocabulary should result in children's lexicons including more words and phrases representing abstract mathematical concepts, while also facilitating more complex understanding of these concepts. Children with larger oral vocabularies should more easily understand and solve mathematics problems presented in spoken or written stories, in part because of a better understanding of abstract terminology (Davidse, De Jong, & Bus, 2014). A relation between children's oral vocabularies and mathematics achievement may emerge prior to school entry (Purpura, Hume, Sims, & Lonigan, 2011).

A larger oral vocabulary should also result in greater behavioral functioning (e.g., Qi & Kaiser, 2004). Developmentally, this should occur as children begin using words to monitor and modify their own behavior (Eisenberg, Sadovsky, & Spinrad, 2005). Young children should begin to adopt their caregiver's regulatory speech, which increasingly emphasizes proactive or inhibitory rather than soothing behaviors (Winsler, Diaz, McCarthy, Atencio, & Chabay, 1999). Having a larger oral vocabulary should provide children with both greater symbolic representations of their internal states and better articulated frameworks for understanding their experiences, resulting in greater ability to organize and guide actions, regulate emotions, and self-verbalize problem-solving strategies (Cole, Armstrong, & Pemberton, 2010). Observable indicators of greater behavioral self-regulation, or "approaches to learning," include remaining attentive, persistent, flexible, engaged, and organized while completing classroom tasks (e.g., Li-Grining, Votruba-Drzal, Maldonado-Carreno, & Haas, 2010). Children's language abilities and behavioral self-regulation positively correlate as early as 24 months of age (Vaughn, Kopp, & Krakow, 1984).

A larger oral vocabulary should also facilitate children's understanding and communication with adults and peers (Cole et al., 2010). Conversely, children with smaller oral vocabularies should display more frequent externalizing or internalizing problem behaviors as they experience greater frustration and anger, lack of control of their environment, peer rejection, task avoidance, and withdrawal (Menting, Van Lier, & Koot, 2010). Having a smaller oral vocabulary positively covaries with greater avoidance and acting out behaviors in children as young as 24 months of age (Rescorla, Ross, & McClure, 2007).

Methodological and Substantive Limitations of Extant Research

Currently, however, the extant work's methodological and substantive limitations constrain empirically derived conclusions as to oral vocabulary's theorized importance for children's academic and behavioral functioning generally and as a specific target of early intervention for at-risk populations.

For example, the National Early Literacy Panel (NELP; 2008) identified 11 factors considered predictive of conventional literacy and concluded that five were "moderately correlated with at least one measure of later literacy achievement but either did not maintain this predictive power when other important contextual variables were accounted for or have not yet been evaluated by researchers in this way" (p. viii). One of these five factors was oral language, including vocabulary. The NELP (2008) suggested that future research evaluate the contribution of these factors.

To date, "surprisingly little" research has been conducted with preschool-aged children, with most studies relying on very small clinical samples (Horwitz et al., 2003, p. 932). Hart and Risley's (1995) seminal study analyzed data from a longitudinal sample of 42 families, including only 6 families representing the lowest socioeconomic status (SES) class. Pan, Rowe, Singer, and Snow (2005) characterized data from these 6 families as constituting "nearly the entirety of what we know" about vocabulary gaps by very young children from low-income families (p. 764). Fernald, Marchman, and Weisleder's (2013) recent study investigating vocabulary gaps at 18 and 24 months of age between children from high- and low-SES families was based on a sample of 48 children. Use of small samples limits generalizability to the larger, heterogeneous U.S. population. This is a limitation of studies analyzing even larger convenience samples (e.g., NICHD Early Child Care Research Network, 2005: Storch & Whitehurst, 2002). Knowledge is also limited about the age of onset and risk factors for vocabulary gaps for population subgroups that, in addition to children from low-SES families (Fernald et al., 2013) and racial or ethnic minorities (Farkas & Beron, 2004), may also be at risk. Associations with other modifiable factors (e.g., parenting, maternal mental health) have yet to be comprehensively examined. Early screening, monitoring, and intervention efforts would be better targeted if guided by findings from large-scale, epidemiological-type studies identifying the sociodemographic, gestational and birth, and family risk and resilience factors most strongly associated with early and meaningful differences in children's oral vocabularies. Many of these factors are known to be associated with or predictive of children's later cognitive, academic, and behavioral functioning (Lynch, 2011). Yet their specific associations with early vocabulary gaps are less clear.

Small convenience samples also limit the field's understanding of whether other factors explain the relation between oral vocabulary size and children's academic and behavioral functioning. For example, whether general cognitive functioning explains the relation between oral vocabulary and children's academic achievement has yet to be firmly established (Rescorla et al., 2007). Although a predictive relation between oral vocabulary size and reading achievement has been repeatedly observed (e.g., NICHD Early Child Care Research Network, 2005; Storch & Whitehurst, 2002; Walker et al., 1994; see Sénéchal, Ouellette, & Rodney, 2006, for a review), as well as comorbidity of language and reading disability (see Pennington & Bishop, 2009, for a review), evidence as to whether the relation is potentially causal is unclear due to the currently limited and mixed evidence (NELP, 2008; Ouellette, 2006; Schatschneider, Fletcher, Francis, Carlson, & Foorman, 2004). To what extent toddler- and preschool-aged children's oral vocabularies predict their later mathematics achievement or, separately, their behavioral functioning has yet to be comprehensively evaluated using population-based data. The extant studies almost entirely investigate single-domain pathways between oral vocabulary and (a) reading achievement, *or* (b) mathematics achievement, *or* (c) behavioral functioning instead of estimating multiple domain pathways simultaneously, particularly as children make the major transition to school (see Appendix S2 for further discussion of existing work's limitations).

Potential Confounds of the Theorized Relation Between Oral Vocabulary and Academic and Behavioral Functioning

Rigorously evaluating the extent to which 24-month-old children's oral vocabularies uniquely predict their academic and behavioral functioning by kindergarten requires extensive statistical control for many potential confounds. Evidence of oral vocabulary's predictive relations following this control would address identified limitations in the field's knowledge base (NELP, 2008), provide stronger support for each relation's potential causality (Hart & Risley, 1995), and empirically evaluate the merits of calls for greater emphasis on oral vocabulary and other language-based competencies in early intervention efforts (Dickinson et al., 2010; Hoff, 2013; Lesaux, 2012). To this end, we analyzed population-based data collected through a non-experimental panel design to evaluate evidence for potential causal relations between 24-month-old children's oral vocabularies and their academic and behavioral functioning at kindergarten entry.

We did so in two ways. First, we investigated whether 24-month-old children's oral vocabularies predicted their academic and behavioral functioning at 60 months of age, establishing temporal precedence. Second, we examined whether 24-month-old children's oral vocabularies continued to predict their academic and behavioral functioning at 60 months of age following extensive statistical control of many factors previously identified as potential confounds (e.g., general cognitive functioning, prior behavioral functioning), reducing the likelihood that any observed predictive relations between earlier oral vocabulary size and later academic and behavioral functioning were spurious.

Potential confounding factors of a relation between 24-month-old children's oral vocabularies and their later academic and behavioral functioning may be grouped into several types of factors, each previously established as predictive of academic and/or behavioral functioning. The first group includes sociodemographic characteristics of families including SES, race or ethnicity, maternal age, and parental marital status. Low SES in particular may result in less cognitively stimulating and higher stress environments that constrain young children's academic and behavioral growth (McLoyd, 1998). Low SES is strongly associated with non-White race or ethnicity, young maternal age, and being a single parent (DeNavas-Walt, Proctor, & Smith, 2013).

The second group of confounds are gestational and birth characteristics, including low or very low birth weight, multiple gestation, and medical risks and complications during pregnancy and at delivery. For instance, being born at low or very low birth weight (for which twins and higher order multiples have heightened risk) may result in neurodevelopmental impairments in behavioral self-regulation (Klebanov, Brooks-Gunn, & McCormick, 1994), general cognitive functioning (Hack, Klein, & Taylor, 1995), and academic achievement (Lynch, 2011). Medical risks during pregnancy (e.g., maternal substance use), and complications of pregnancy (e.g., gestational diabetes) and delivery (e.g., prolonged labor) are associated with later developmental delays (e.g., Anthopolos, Edwards, & Miranda, 2013).

Parenting and the quality of the home environment constitute a third group of potential confounds, with family stress and investment potentially explaining the effects of SES on children's development (Guo & Harris, 2000). For instance, young children whose parents are warm and supportive, set consistent routines for their children, and provide more cognitively stimulating environments often display greater academic and behavioral functioning, even after accounting for SES and other sociodemographic characteristics (Iruka, LaForett, & Odom, 2012). Other potentially important features of the home and family include whether a family member has a mental illness, learning disability, or special need; whether the mother has health problems, is depressed, or socially isolated; how much time the child spends watching television; and whether or not the child attends day care. Parental mental and physical health problems are associated with cognitive and other delays in young children (Breaux, Harvey, & Lugo-Candelas, 2013). Toddlers and preschool-aged children who frequently watch television are more likely to enter school with lower academic functioning, particularly in reading (Pagani, Fitzpatrick, & Barnett, 2013). This is possibly due to television viewing displacing storybook reading (Koolstra & Van der Voort, 1996) and limiting growth in children's cognitive and attentional capacities (Christakis, Zimmerman, DiGiuseppe, & McCarty, 2004). Conversely, educational or subtitled programs may instead help increase young children's achievement (Ennemoser & Schneider, 2007). Frequently attending child care may increase children's risk for behavior problems, including during the transition to kindergarten (NICHD Early Child Care Research Network, 2003).

Young children's own level of cognitive and behavioral functioning prior to or by kindergarten entry constitute additional potential confounds. Cognitive functioning strongly predicts children's later academic achievement and behavior (Duncan et al., 2007), and is likely related to children's oral vocabularies (e.g., Menting et al., 2010). Behavioral self-regulation (e.g., attentiveness, task persistence), externalizing behavior problems (e.g., aggressiveness), and internalizing behavior problems (e.g., anxiety, withdrawal) are also associated with children's oral vocabularies (Menting et al., 2010) and should autoregressively predict later behavioral functioning (Morgan, Farkas, & Wu, 2009).

Study's Purpose

We sought to identify sociodemographic, gestational and birth, cognitive and behavioral, and family functioning factors associated with or predictive of U.S. children's oral vocabularies at 24 months of age. We also evaluated whether and to what extent the children's oral vocabularies uniquely predicted their academic and behavioral functioning at kindergarten entry. Our study investigated three specific research questions. First, which U.S. children display larger or smaller oral vocabularies at 24 months of age? Second, does having a larger oral vocabulary at 24 months of age uniquely predict greater academic functioning at kindergarten entry? Third, does having a larger oral vocabulary at 24 months of age uniquely predict greater behavioral functioning at kindergarten entry?

Method

Design and Analytical Samples

We analyzed data from the Early Childhood Longitudinal Study–Birth Cohort (ECLS–B), a population-based, longitudinal cohort assessed from birth to kindergarten entry (http://nces.ed.gov/ecls/birth.asp). This sample was selected from birth certificate records and includes oversamples of particular population subgroups (e.g., Native Americans and Alaska Natives, children born with very low birth weight), with sample weights provided to generate nationally representative estimates. ECLS–B field staff individually administered measures of children's cognitive and academic functioning and conducted interviews with family members, when the children were 9, 24, 48, and 60 months of age. Kindergarten teachers rated the children's behaviors at school entry.

We identified three analytical samples. For the 0- to 24-month analyses, children with missing data on the study's 24-month oral vocabulary measure were excluded, as were children with congenital anomalies. We used multiple imputation (MI) procedures to account for missing data for predictor variables but not for the oral vocabulary variable in the remaining cases. Specifically, we used Imputation and Variance Estimation Software or IVEware (http://www.isr.umich.edu/src/smp/ive/) to repeatedly replace missing values with predictions based on random draws from the posterior distributions of observed sample parameters, which results in multiple complete data sets (Raghunathan, Solenberger, & Van Hoewyk, 2002). We averaged results obtained across five different imputation samples to account for random variations in the data sets derived from MI (Raghunathan et al., 2002). This resulted in an analytical sample of 8,650 with data for analyses evaluating factors predictive of 24-month-old children's oral vocabularies. For the 24- to 60-month analyses of reading and mathematics achievement, we again multiply imputed the predictor variables but excluded cases with missing data on 60-month reading or mathematics achievement measures ($N = 6{,}050$). For the 24- to 60-month analyses of behavioral functioning, we multiply imputed the predictor variables but excluded cases with missing data on the behavioral measures ($N = 4{,}350$). (All reported sample sizes are rounded to the nearest 50 to comply with National Center for Education Statistics [NCES], U.S. Department of Education requirements for data reporting.) Table 1 displays weighted percentages or means or standard deviations for the variables. The means of the variables are quite similar across the three samples, suggesting that the samples are comparable despite varying in size due to missing dependent variable data.

Measures, Criterion Variables

Oral Vocabulary, 24 Months

Oral vocabulary size at 24 months was assessed using a modified version of the MacArthur–Communicative Development Inventory (M–CDI; Fenson, Dale, Reznick, Thal, & Bates, 1993) developed specifically for the ECLS–B by the CDI Advisory Board. This measure is a parent survey of children's expressive use of 50 words commonly known and spoken in the targeted age range. An equivalent list of Spanish words was provided for Spanish-speaking parents; non-Spanish-speaking language minority parents were not included in the analyses. Internal consistency for the original M–CDI vocabulary scales is high ($\alpha = .96$). The M–CDI displays construct, concurrent, and predictive validity (Fenson et al., 1994) and classifies children into language status groups with 97% accuracy (Skarakis-Doyle, Campbell, & Dempsey, 2009). The number of words spoken by children as identified by parents was summed to create the Total Word score at 24 months. The mean of this variable was about 29 words, with a standard deviation of about 12. (Because this score was obtained using the shortened version of the M–CDI, it is not directly comparable to the 24-month mean number of words spoken measured by the original M–CDI.)

Reading and Mathematics Achievement, 60 Months

The ECLS–B Reading Test consisted of 55 items designed to assess language development, emergent literacy, and basic reading. The Mathematics Test included 42 items designed to assess number sense, counting, operations, geometry, patterns, and measurement. Each measure consisted of a two-stage routing procedure and item response theory (IRT)

Table 1
Weighted Percentages or Means with Standard Deviations of the Study's Three Analytical Samples

Variables	Sample 1[a]	Sample 2[b]	Sample 3[c]
Total Word score			
24 months	28.57 (11.83)	28.84 (11.79)	28.72 (12.00)
Reading Test score			
60 months	38.39 (14.83)	38.35 (14.75)	38.55 (14.86)
Mathematics Test score			
60 months	40.17 (10.56)	40.24 (10.53)	40.49 (10.38)
Approaches to Learning			
24 months	13.92 (3.56)	13.98 (3.53)	13.97 (3.59)
60 months	18.53 (2.49)	18.51 (2.50)	18.50 (2.50)
Internalizing Problems			
24 months	4.41 (1.83)	4.35 (1.80)	4.40 (1.83)
60 months	6.31 (2.19)	6.33 (2.20)	6.32 (2.20)
Externalizing Problems			
24 months	4.87 (1.75)	4.84 (1.73)	4.86 (1.76)
60 months	6.96 (3.07)	7.01 (3.10)	7.02 (3.12)
Race			
White	54.2%	53.9%	54.4%
Black	13.6%	14.0%	13.0%
Hispanic	24.8%	25.0%	25.5%
Asian	2.7%	2.6%	2.6%
Native American	0.5%	0.5%	0.5%
Other	4.2%	4.1%	4.0%
Age			
24 months	24.16 (0.70)	24.15 (0.69)	24.15 (0.68)
Gender			
Male	51.1%	50.6%	51.5%
SES quintiles			
Lowest	19.3%	19.1%	18.4%
Second lowest	19.6%	20.0%	20.2%
Middle	20.0%	20.6%	21.0%
Second highest	20.7%	20.9%	20.8%
Highest	20.3%	19.4%	19.6%
Nonsingleton	3.3%	3.2%	3.3%
Mother's age at child's birth			
Equal to or older than 35	14.0%	13.7%	13.5%
Younger or equal to 18	6.9%	7.3%	7.4%
Marital status			
Not married	31.4%	31.2%	31.4%
Primary language			
Non-English	18.6%	18.0%	18.1%
Birth weight			
Very low	1.2%	1.2%	1.2%
Moderately low	6.2%	6.2%	6.2%
Labor complications	35.5%	34.8%	35.0%
Obstetric procedures	58.5%	57.5%	56.7%
Risks			
Medical	18.1%	17.7%	17.9%
Behavioral	11.1%	10.8%	10.7%
Parenting score			
24 months	7.32 (1.15)	7.33 (1.14)	7.34 (1.14)
Family member status			
Mental illness	10.7%	11.1%	10.7%
Learning disability	15.3%	15.4%	15.7%
Maternal status			
Health problems	7.3%	7.2%	7.0%
Household status			
Special need	7.4%	7.9%	7.8%
Mother depressed	13.7%	13.9%	14.7%
Stayed in day care center			
> 10 hr per week	14.0%	13.5%	13.8%
Child television usage			
Middle third	34.4%	33.9%	34.5%
Highest third	32.6%	33.0%	32.5%
Mother isolated	0.16 (0.44)	0.15 (0.44)	0.15 (0.45)
Bayley Mental score			
24 months	126.60 (10.49)	126.86 (10.33)	127.11 (10.67)

Note. Sample size rounded to nearest 50 per ECLS–B confidentiality requirements. Sample 1: N = 8,650; Sample 2: N = 6,050; Sample 3: N = 4,350. ECLS–B = Early Childhood Longitudinal Study-Birth Cohort; SES = socioeconomic status.
[a]Sample of ECLS–B children aged 0–24 months of age with complete data on Total Word score measure at 24 months of age. [b]Sample of ECLS–B children aged 24–60 months of age with complete data on the reading and mathematics achievement measures administered at 60 months of age. [c]Sample of ECLS–B children aged 24–60 months of age with complete data on the approaches to learning, externalizing problem behaviors, and internalizing problem behaviors measures administered at 60 months of age.

scaling. All children were initially given the same 24-question test. Then, and depending on the number of correct responses on this test, they were administered one of three follow-up routing test forms on the basis of scoring in the low, middle, or high range of the initial test. The Reading and the Mathematics Tests display IRT theta reliability coefficients of .92, indicating high reliability of the assessment scores (Najarian, Snow, Lennon, & Kinsey, 2010).

Behavioral Self-Regulation, Externalizing Problem Behaviors, and Internalizing Problem Behaviors, 60 Months

Teachers rated children's behavioral functioning using items from the Preschool Learning and Behavior Scales (2nd ed.), the Social Skills Rating

System, and the Early Childhood Longitudinal Study, Kindergarten Class of 1998–1999. Teachers rated the children's frequency of behaviors using a scale ranging from 1 (*never*) to 5 (*very often*). We conducted an exploratory factor analysis, using a promax rotation, of 22-rated behaviors, retaining a four-factor solution after examining several potential factor solutions and considering a priori criteria. Items with factor loadings lower than .60 were removed. We identified three of the four factors as relevant to this study: (a) approaches to learning, (b) internalizing problem behaviors, and (c) externalizing problem behaviors. The Approaches to Learning scale (α = .91) contains five items measuring behavioral self-regulation (e.g., "pays attention well," "keeps working until finished," and [reverse scaled] "has difficulty concentrating"). The Internalizing Problem Behaviors scale (α = .64) contains three items related to anxious or withdrawn behaviors (i.e. "seems unhappy," "worries about things," and "acts shy"). The Externalizing Problem Behavior scale (α = .87) contains four items related to acting-out behaviors (e.g., "disrupts others," "has temper tantrums," "is physically aggressive"). Items were summed to obtain scale scores. Higher Approaches to Learning scale scores indicated more appropriate classroom-based behavioral self-regulation; higher Internalizing or Externalizing Problem Behaviors scale scores indicated more frequent problem behaviors.

Measures, Predictor Variables

Sociodemographic Characteristics

Sociodemographic data were collected in parent interviews at each assessment, as well as from birth certificates. Race or ethnicity was defined as the race or ethnicity of the mother from birth certificates. Non-Hispanic White was the reference group; other groups included non-Hispanic Black, Hispanic, Asian, Native American, and Other. Child age in months was included to account for variations in actual age at the time of assessment. Female was the reference gender category. ECLS–B staff calculated family SES using a composite of five parent-reported indicators that included the father/male guardian's education and occupation, the mother/female guardian's education and occupation, and household income. NCES estimated missing values, which averaged 2.5% across the five SES indicators, using hot deck imputation. The SES distribution was divided into quintiles and is represented in regressions as a set of dummy variables. The first quintile represented the lowest SES. Children born individually were considered singletons; all others were coded as nonsingletons. Dummy variables for whether the mother was equal to or over 35 years of age or less than or equal to 18 years of age at the time of the child's birth were also included. Being married was used as the reference category for marital status. Children living in households where English was not the primary language spoken were identified using a dichotomous variable.

Characteristics From Birth Certificates

A count of medical conditions in pregnancy included incompetent cervix, acute or chronic lung disease, chronic hypertension, pregnancy-induced hypertension, eclampsia, diabetes, hemoglobinopathy, cardiac disease, anemia, renal disease, genital herpes, oligohydramnios, uterine bleeding, Rh sensitization, previous birth weighing 4,000+ g, or previous preterm birth. Behavioral risk factors during pregnancy included any maternal use of alcohol and/or tobacco during pregnancy (coded as 1 if present and summed to form a scale that ranged from 0 to 2). Obstetrical procedures were measured as a count of total procedures including induction of labor, stimulation of labor, tocolysis, amniocentesis, and Cesarean section. Labor complications were measured as a count of the following: placental abruption anesthetic complications, dysfunctional labor, breech/malpresentation, cephalopelvic disproportion, cord prolapse, fetal distress, excessive bleeding, fever of > 100°F, moderate/heavy meconium, precipitous labor (< 3 hr), prolonged labor (> 24 hr), placental previa or seizures during labor. Two indicator variables were used to quantify very low (<1,500 g) and moderately low (1,500–2,500 g) birth weight. Normal (>2,500 g) birth weight was the reference category.

Parenting Quality, 24 Months

We averaged scores on two assessments conducted at 24 months to create a composite parenting score. The first assessment included items from the Home Observation for Measurement of the Environment (HOME; Bradley & Caldwell, 1984), a widely used measure of the quality of the child's parenting and the home environment (e.g., NICHD Early Childcare Research Network, 2005). NCES retained a subset of the original measure's 21 items measuring (a) parental activities including reading to the child, telling stories, singing, and taking the child on errands or to public places; (b) having toys, records, books, and audiotapes available in the home; and (c) having a safe and supportive home environment. Because the modified HOME scale had relatively low internal consistency ($\alpha = .46$), the ECLS–B manual advises researchers to use alternatives other than scaling the items. We therefore used a count of 14 HOME scores items to indicate the extent to which these positive activities were reported or observed.

The second parenting assessment consisted of ratings of the quality of a parent's interactive support of their child, as coded from videotaped interactions during the Two Bags Task, a simplified version of the Three Bags Task used in the Early Head Start Research and Evaluation Project and the NICHD Study of Early Child Care (Nord, Edwards, Andreassen, Green, & Wallner-Allen, 2006). Interviewers read a script to the children's parents, after which, over the next 10 min, parents were asked to play with their children using a wordless picture book (i.e., *Goodnight Gorilla*) and a set of toy dishes. A composite variable measuring parental support was created for the ECLS–B representing the mean of three characteristics of parental interactions with their children: (a) sensitivity, (b) stimulation of cognitive development, and (c) positive regard. Each item was scored on a 7-point scale, ranging from 1 = *very low* to 7 = *very high*. Mean interrater reliability for the parent rating scales was 97%, with mean reliabilities of 97%, 93%, and 94% for sensitivity, cognitive stimulation, and positive regard, respectively (see Andreassen, Fletcher, & Park, 2007, for additional measurement detail).

History of Mental or Physical Illness or Disabling Conditions, 9 Months

Family member with mental illness. Mothers were asked at the 9-month survey whether they or a family member had "a serious mental illness, such as schizophrenia, a paranoid disorder, a bipolar disorder, or manic episodes?" "Yes" answers were coded as 1; "no" answers were coded as 0.

Family member with learning disability. Mothers were also asked at 9 months, "Have you or any of your blood relatives ever had a learning disability?" "Yes" answers were coded as 1; "no" answers were coded as 0.

Maternal health problems. During the 9-month wave, mothers rated their overall health on a scale of 1–5. The responses were transformed into a dichotomous variable, with responses of 4 (*fair*) or 5 (*poor*) coded as 1. All other responses (e.g., "excellent") were coded as 0.

Special needs. Mothers responded during the 9-month survey whether they or any other household members had a special need, delay or disability. "Yes" responses were coded as 1; "no" responses were coded as 0.

Maternal depression. A modified version of the Center for Epidemiologic Studies–Depression (CES–D) scale (Radloff, 1977) was used as part of the 9-month parent self-administered questionnaire. The modified CES–D scale includes 12 items including having poor appetite, feeling lonely, and trouble keeping focus. Scores were dichotomized so that a total score greater than 9 was coded as 1 for presence of depressive symptoms, corresponding to the cutoff commonly used for the full CES–D of greater than 15 (Nord et al., 2006).

Maternal social isolation. Five variables were used to construct a scale of maternal social isolation. Four of these variables asked the mother if she had people available to ask for help for various needs (1 = "no one" for each variable). The other variable asked about being close to her own mother (1 = "not close"). The sum of these five variables was calculated and used as a measure of social isolation.

Hours per Week in Child-Care Center, 24 Months

During the 24-month survey, parents were asked how many hours a week their child spends in a child-care center. Responses to this question were coded dichotomously so that a response of 11 hr or more was coded as 1.

Television Usage, 24 Months

Average weekly television usage was calculated based on a parent's response at 24 months about the child's average number of hours spent watching television or videos in the household during weekdays and nends. Less than 9 hr was considered low television usage, between 9 and 17 hr was considered medium television usage, and more than 17 hr was considered high television usage.

Cognitive Functioning, 24 Months

A standardized assessment was administered at the 24-month survey wave to measure children's general cognitive functioning. Field staff individually administered the Bayley Short Form–Research Edition (BSF–R), a modified version of the Bayley

Scales of Infant Development, 2nd ed. (BSID–II; Bayley, 1993), which is designed for use with children from birth to 36 months of age. In both the BSID–II and the BSF–R, the mental score is based on the trained interviewer's assessment of a child's ability to perform tasks related to memory, habituation, preverbal communication, problem solving, and concept attainment. The IRT reliability coefficient of the BSF–R mental scale at 24 months was .88 (Andreassen et al., 2007). The reported R^2 between BSF–R and BSID–II scores was .99, indicating that the BSF–R maintains the psychometric properties of the BSID–II and accurately measures children's performance across the ability distribution (Andreassen et al., 2007).

Behavioral Self-Regulation, Externalizing Problem Behaviors, and Internalizing Problem Behaviors, 24 Months

Early Childhood Longitudinal Study–Birth Cohort trained examiners used the Behavior Rating Scale–Research Edition (BRS–R) to rate children's behaviors as they completed BSF–R tasks during the 24-month survey wave. The BRS–R was adapted from the Behavior Rating Scale (BRS; Bayley, 1993). Scores on the BRS correlate moderately to highly with scores on other measures of young children's socioemotional adjustment (Buck, 1997). The BRS–R included 11 interviewer-rated items from the full BRS measuring developmentally appropriate behaviors for 24-month-old children including their attention to task, persistence, cooperation with an examiner, interest in the testing materials, and frustration with testing tasks (Nord et al., 2006). A 5-point scale was used to rate the frequency of the observed behavior. A higher score indicated that the problem behavior occurred infrequently. The self-regulatory items on the BRS (e.g., attention to task, persistence) had an internal consistency of .92.

We used eight items from the BRS–R to control for prior behavioral functioning. Four items measured children's prior behavioral self-regulation (e.g., "pays attention to tasks," "is persistent in tasks"). Two items measured prior externalizing behaviors (i.e., "frustration in tasks," "cooperation"). Two items measured prior internalizing problem behaviors (i.e., "fearlessness," "social engagement"). Specific items were reverse coded as appropriate to be consistent with either appropriate (i.e., for behavioral self-regulation) or problematic (i.e., for externalizing or internalizing) behavioral functioning. Cronbach's alphas were .90, .64, and .72 for the 24-month behavioral self-regulation, externalizing problem behaviors, and internalizing problem behaviors measures, respectively.

Analytical Methods

The score distribution of the measure of 24-month-old children's parent-reported oral vocabulary generally resembled a normal curve, with a truncation point at 50 words (i.e., the maximum possible score). There are 280 cases of about 9,500 (i.e., 2.9%) at this truncation point. To avoid bias due to truncation at the upper limit of the variable (ceiling effect), we used Tobit regression analysis when predicting children's scores on the oral vocabulary measure. Our first set of analyses estimated six Tobit regression models predicting 24-month-old children's oral vocabularies. We entered these sets of predictors sequentially. As shown in Table 2, Model 1 estimated to what extent children's or families' sociodemographic characteristics functioned as predictors. The sociodemographic variables included children's race or ethnicity, age, gender, and SES. Model 2 added additional sociodemographic variables to the regression equation. Model 3 added the children's gestational and birth characteristics. Model 4 entered predictors related to family risk and resilience. Model 5 added children's behavioral functioning; Model 6 added their cognitive functioning. Estimating these models sequentially allowed us to investigate whether there were significant direct effects of these groupings of variables on 24-month-old children's oral vocabularies, as well as whether these groupings of variables explained the direct effects of the previously entered variables.

Our second and third sets of analyses investigated whether and to what extent, and before and after extensive statistical control, having a larger oral vocabulary at 24 months of age predicted children's academic and behavioral functioning at 60 months of age. We conducted separate ordinary least squares regression analyses predicting each of the five criterion variables. Predictor variables in these regression analyses included oral vocabulary, sociodemographic and birth characteristics, family risk and resilience characteristics, and children's cognitive and behavioral functioning at 24 months of age. We used SAS 9.3 (SAS Institute Inc., Cary, NC) to perform the analyses. We incorporated sampling weights and design effects to account for oversampling of some population subgroups and for the stratified cluster design of the ECLS–B. We used an alpha of $p < .05$ in establishing statistical significance including for our primary analyses

Table 2

Parameter Estimates (Standardized) of Tobit Regression Equations Predicting Children's Oral Vocabularies at 24 Months, ECLS–B Data

Variables	Model 1[a]	Model 2[b]	Model 3[c]	Model 4[d]	Model 5[e]	Model 6[f]
Intercept	0.92***	0.98***	0.96***	0.75***	0.69***	0.26***
Race						
Black	−0.02	0.01	0.03	0.17***	0.16***	0.25***
Hispanic	−0.17***	−0.12**	−0.12**	−0.06	−0.06	0.05
Asian	−0.13	−0.05	−0.05	0.12	0.13	0.11
Native American	−0.15	−0.16	−0.16	−0.07	−0.04	0.16
Other	−0.06	−0.08	−0.06	−0.0001	0.02	0.08
Age						
24 months	0.19***	0.18***	0.18***	0.18***	0.17***	0.08***
Gender						
Male	−0.52***	−0.51***	−0.52***	−0.49***	−0.44***	−0.26***
SES quintile						
Lowest	−0.55***	−0.57***	−0.54***	−0.22***	−0.21***	−0.08*
Second lowest	−0.44***	−0.47***	−0.45***	−0.20***	−0.22***	−0.06
Middle	−0.32***	−0.34***	−0.33***	−0.17***	−0.18***	−0.04
Second highest	−0.21***	−0.22***	−0.22***	−0.13**	−0.12**	−0.02
Nonsingleton		−0.47***	−0.30***	−0.28***	−0.24***	−0.13*
Mother's age at child's birth						
Equal to or older than 35		−0.13***	−0.13**	−0.14***	−0.14***	−0.11***
Younger or equal to 18		−0.05	−0.04	−0.03	−0.01	−0.05
Marital status						
Not married		0.02	0.03	0.06	0.08*	0.08**
Primary language						
Non-English		−0.12*	−0.12**	−0.05	−0.05	0.10*
Birth weight						
Very low			−0.88***	−0.87***	−0.83***	−0.39***
Moderately low			−0.24***	−0.23***	−0.21***	−0.05
Labor complications			0.01	0.002	0.01	0.001
Obstetric procedures			0.05**	0.05*	0.05*	0.04**
Risks						
Medical			−0.04	−0.03	0.01	−0.03
Behavioral			−0.06	−0.04	−0.01	0.01
Parenting score at 24 months				0.28***	0.24***	0.10***
Family member status						
Mental illness				−0.12**	−0.12**	−0.09
Learning disability				−0.03	−0.02	0.03
Maternal status						
Health problems				−0.27***	−0.24***	−0.15***
Household status						
Special need				−0.16***	−0.14**	−0.05
Mother depressed				0.02	0.03	0.02
Stayed in day-care center > 10 hr per week				0.12**	0.12**	0.03
Child television usage						
Middle third				0.01	0.10	−0.01
Highest third				−0.11***	−0.09**	−0.04
Mother isolated				−0.03*	−0.03*	−0.02
Approaches to learning 24 months					0.62***	0.20***
Internalizing problems 24 months					−0.27***	−0.08**

Table 2
Continued

Variables	Model 1[a]	Model 2[b]	Model 3[c]	Model 4[d]	Model 5[e]	Model 6[f]
Externalizing problems						
24 months					−0.30***	−0.07
General cognitive functioning						
24 months						0.61***

Note. $N = 8{,}650$. Sample size rounded to nearest 50 per ECLS–B confidentiality requirements. ECLS–B = Early Childhood Longitudinal Study–Birth Cohort; SES = socioeconomic status.
[a]Model 1 = Intercept, race, age, gender, and SES. [b]Model 2 = Adds maternal age, marital status, and primary language spoken in the home as additional predictors. [c]Model 3 = Adds birth weight, labor complications, obstetric procedures, medical and behavioral risks as predictors. [d]Adds parenting, family and maternal and household status, maternal depression, child care, television usage, and maternal isolation as predictors. [e]Adds behavioral self-regulation and problem behaviors as predictors. [f]Adds general cognitive functioning as a predictor.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

(Moyé, 2008; see Appendix S3 for multiplicity check information).

Results

Which U.S. Children Display Larger or Smaller Oral Vocabularies at 24 Months of Age?

Table 2 displays results from Tobit regressions predicting children's oral vocabularies at 24 months of age. All continuous variables (including the dependent variable) were standardized. Thus, all coefficients represent directly comparable standardized coefficients (i.e., effect sizes).

The first column (Model 1) shows the very strong and significant effects of gender and SES. Boys averaged .52 *SD* fewer words spoken than girls. This gap is similar in magnitude to the vocabulary gap by U.S. children in the lowest and highest SES quintiles. (Thus, a boy toddler from the lowest SES quintile, on average, experienced a double decrement summing to somewhat more than 1 *SD*.) The association between SES quintiles and words spoken was approximately linear, with a difference of about .10 *SD* for each successive quintile. In contrast, there was only an inconsistent association between children's race or ethnicity and oral vocabularies.

The second column shows that, among additional sociodemographics, the largest association with 24-month-old children's oral vocabularies involved the .47 *SD* decrement experienced by nonsingletons. Although this association was reduced in size in subsequent models, it remained significant. A modest (.13 *SD*) but significant decrement was observed for the children of older mothers. A decrement of .12 was observed for children from homes where English was not the primary spoken language. Controlling for this language variable also partially explained the negative relation between children's Hispanic status and their oral vocabulary size.

The third column adds gestational and birth characteristics as predictors. There was a very strong negative predictive relation (−.88 *SD*) between very low birth weight and 24-month-old children's oral vocabularies. The relation with moderately low birth weight was smaller (−.24 *SD*) but also statistically significant. The effects of labor complications, obstetric procedures, and the mother's medical and behavioral risks during pregnancy were small. Including these covariates in the regression equation reduced the predicted effect of being a nonsingleton by about one third. Thus, a portion of this relation appeared to be due to the relatively lower birth weights of nonsingletons. The fourth column displays the results of adding family risk and resilience factors to the regression equation. Two resilience factors—parenting quality and day-care attendance—were both significantly associated with 24-month-old children having larger oral vocabularies, with predicted effect sizes of .28 and .12 *SD*, respectively. Among the risk factors, several were also significantly and negatively associated with oral vocabulary. These included a family member with a mental illness (−.12 *SD*), maternal health problems (−.27 *SD*), a household member with special needs (−.16 *SD*), and the child being in the highest third of television usage (−.11 *SD*). These factors helped explain the predicted effect of SES. Following their statistical control, the association between each SES quintile and children's oral vocabularies decreased by about 50% but remained significant.

The fifth and sixth columns of Table 2 show the results of adding the 24-month behavioral and general cognitive functioning measures into the regression equation. The three behavioral measures were associated in expected directions with oral vocabulary, so that 24-month-old children who displayed more attentive, task persistent, and other types of learning-related behaviors also displayed larger oral vocabularies (.62 *SD*), while those children with stronger internalizing or externalizing problem behaviors displayed smaller oral vocabularies (−.27 and −.30 *SD*, respectively).

Because the measures of general cognitive functioning and oral vocabulary were both language dependent, we expected to find a strong association between the two variables. As shown in Model 6, the relation was indeed very strong and significant (.61 *SD*). However, even with these behavioral and general cognitive functioning controls added to the regression equation, most of the associations remained statistically significant despite decreasing in magnitude. The exception was family SES, suggesting that the relation between SES and 24-month-old children's oral vocabularies was partially accounted for by other risk factors. These included being a raised by a single mother, being born with low birth weight, experiencing less warm or cognitively stimulating parenting, being raised in a household with a socially isolated mother or one who has health problems, and, mostly strongly, by children's own level of general cognitive and behavioral functioning. However, being in the lowest SES quintile remained significantly negatively related to children's oral vocabulary size despite extensive statistical control. Oral vocabulary at 24 months functioned somewhat independently of general cognitive functioning, which is a conclusion that is also supported by results in the study's second set of analyses.

Does Having a Larger Oral Vocabulary at 24 Months of Age Uniquely Predict Greater Academic Functioning at Kindergarten Entry?

Table 3 displays the results of using 24-month-old children's oral vocabulary, as well as the full set of additional risk and resilience factors just presented, to predict their academic functioning at kindergarten entry. The first model of Table 3 displays the unadjusted estimate between oral vocabulary and reading or mathematics achievement. Then the second model adds control variables to the regression equation. The third column adds general cognitive functioning as a control, allowing for a highly conservative test of the hypothesized relation. Oral vocabulary size initially predicted 24-month-old children's reading and mathematics achievement at kindergarten entry. These predicted effect sizes were .22 and .27 *SD*, respectively.

Model 2 shows that despite statistical control for a wide range of covariates, having a larger oral vocabulary at 24 months of age remained positively and significantly predictive of greater academic functioning at kindergarten entry. Model 3's results indicate that even after controlling for 24-month-old children's general cognitive functioning, having a larger oral vocabulary positively and significantly predicted their later reading and mathematics achievement. The adjusted effect sizes were .07 and .10 *SD* for reading and mathematics achievement, respectively.

Table 3 identifies additional factors consistently predictive of greater academic functioning. These included being White or Asian, being from a high-SES family, being born as a singleton, watching television more frequently, and displaying higher levels of general cognitive functioning. However, 24-month-old children's oral vocabularies remained uniquely predictive of their academic functioning at kindergarten entry despite the large number of controls. We also investigated (results not shown) the possibility that SES moderated the predicted effect of oral vocabulary on academic functioning (as well as behavioral functioning). We did so by adding interaction terms between the four SES dummy variables and oral vocabulary to the regression equations. These interaction terms were not statistically significant.

Does Having a Larger Oral Vocabulary at 24 Months of Age Uniquely Predict Greater Behavioral Functioning at Kindergarten Entry?

Table 4 displays regressions predicting teacher ratings of kindergarten children's behavioral self-regulation, as well as their internalizing and externalizing problem behaviors. Having a larger oral vocabulary significantly predicted all three indicators of behavioral functioning in Model 1. Specifically, 24-month-old children with larger oral vocabularies displayed greater behavioral self-regulation (.22 *SD*) and fewer internalizing and externalizing problem behaviors at kindergarten entry (−.11 and −.14 *SD*, respectively). Adding Model 2 and 3's controls to the regression equation reduced oral vocabulary's predictive relations with the three indicators of children's behavioral functioning but the relations remained statistically significant. Having a larger oral vocabulary predicted greater

Table 3

Parameter Estimates (Standardized) of OLS Regressions Predicting Children's Academic Achievement at 60 Months, ECLS–B Data

	Reading achievement			Mathematics achievement		
Variables	Model 1^{a}	Model 2^{b}	Model 3^{c}	Model 1^{a}	Model 2^{b}	Model 3^{c}
Intercept	−0.07**	0.37***	0.33***	−0.05*	0.38***	0.31***
Oral vocabulary						
24 months	0.22***	0.12***	0.07***	0.27***	0.17***	0.10***
Race						
Black		−0.02	0.003		−0.20***	−0.17**
Hispanic		−0.04	−0.02		−0.14*	−0.11
Asian		0.46***	0.47***		0.30***	0.31***
Native American		−0.19*	−0.15		−0.36***	−0.30**
Other		−0.08	−0.06		−0.14	−0.12
Child age						
24 months		0.14***	0.13***		0.12***	0.11***
Gender						
Male		−0.10**	−0.08*		0.01	0.03
SES quintile						
Lowest		−0.78***	−0.76***		−0.77***	−0.75***
Second lowest		−0.59***	−0.57***		−0.59***	−0.57***
Middle		−0.39***	−0.37***		−0.40***	−0.38***
Second highest		−0.24***	−0.23***		−0.23***	−0.22***
Nonsingleton		−0.08*	−0.08*		−0.10*	−0.09*
Mother's age at child's birth						
Equal to or older than 35		−0.01	−0.02		0.03	0.02
Younger or equal to 18		−0.10	−0.11		−0.11*	−0.12*
Marital status						
Not married		−0.07	−0.07		−0.04	−0.04
Primary language						
Non-English		−0.03	−0.004		−0.002	0.03
Birth weight						
Very low		−0.10*	−0.05		−0.28***	−0.21***
Moderately low		−0.03	−0.01		−0.07	−0.04
Labor complications		−0.06	−0.06		−0.003	−0.003
Obstetric procedures		0.02	0.02		0.01	0.02
Risks						
Medical		−0.07*	−0.07*		−0.004	−0.01
Behavioral		−0.06	−0.05		−0.06	−0.04
Parenting score						
24 months		0.03	0.02		0.05*	0.02
Family member status						
Mental illness		0.04	0.05		0.01	0.01
Learning disability		−0.07	−0.07		−0.10*	−0.09*
Maternal status						
Health problems		−0.03	−0.02		−0.08	−0.07
Household status						
Special need		0.02	0.02		−0.02	−0.01
Mother depressed		−0.02	−0.03		−0.04	−0.05
Stayed in day-care center > 10 hr per week		0.04	0.03		0.05	0.04
Child television usage						
Middle third		0.10*	0.09*		0.11**	0.10**
Highest third		0.10*	0.11**		0.11**	0.12**
Mother isolated		−0.02	−0.02		−0.05***	−0.05**
Approaches to learning						
24 months		0.18**	0.12*		0.14**	0.06

Table 3
Continued

	Reading achievement			Mathematics achievement		
Variables	Model 1[a]	Model 2[b]	Model 3[c]	Model 1[a]	Model 2[b]	Model 3[c]
Internalizing problems						
24 months		0.04	0.07		0.02	0.05
Externalizing problems						
24 months		−0.05	−0.02		−0.06	−0.02
General cognitive functioning						
24 months			0.12***			0.16***

Note. $N = 6{,}050$. Sample size rounded to nearest 50 per ECLS–B confidentiality requirements. OLS = ordinary least squares; ECLS–B = Early Childhood Longitudinal Study–Birth Cohort; SES = socioeconomic status.
[a] Intercept and oral vocabulary as predictors. [b] Adds race, age, gender, SES, nonsingleton, maternal age, language spoken in the home, birth weight, labor complications, obstetric procedures, medical and behavioral risks, parenting, family and maternal and household status, maternal depression, child care, television usage, maternal isolation, behavioral self-regulation and problem behaviors as predictors. [c] Adds general cognitive functioning as a predictor.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

behavioral self-regulation (.10 *SD*), as well as fewer internalizing (−.09 *SD*) and externalizing problem behaviors (−.06 *SD*) even following statistical control for autoregressive behavioral functioning. Other less consistent predictors of children's behavioral functioning included being a boy, being raised by a single mother, and frequently attending child care.

Discussion

Our analyses of a population-based data set identify which groups of 24-month-old children in the United States are at risk of having smaller oral vocabularies. Our results are consistent with prior research indicating that children being raised in low-SES families are likely to have smaller oral vocabularies (e.g., Fernald et al., 2013; Hart & Risley, 1995). This specific relation is partially explained by a more general set of sociodemographic, gestational and birth, family risk and resilience, and individual characteristics, particularly children's own levels of cognitive functioning. Other studies report that family stress and investment characteristics mediate the relation between lower SES and academic achievement, but the extent to which this occurs for oral vocabulary has been unclear (Farkas & Beron, 2004), in part due to a range of measures, sampling, and analytical methods being used when examining SES (Hoff, Laursen, & Bridges, 2012). Our results indicate that family stress and investment characteristics partially explain the relation between family SES and children's oral vocabulary size.

The results further indicate that 24-month-old children with smaller oral vocabularies are at risk of entering kindergarten with lower academic and behavioral functioning. Consistent with previously theorized mechanisms, oral vocabulary has both a general and unique relation with children's development. The relation is evident across multiple indicators of both academic and behavioral functioning during early childhood. Twenty-four-month children's oral vocabularies remain predictive of their academic and behavioral functioning at kindergarten entry despite extensive statistical control for many factors previously identified as potential confounds.

Limitations

This study has at least five limitations. First, our measure of children's oral vocabulary size is brief and relies on parental recall. We also used children's scores on this measure to predict their performance on independently administered measures of academic and behavioral functioning 3 years later. The resulting measurement error, as well as multiyear time interval, should bias our estimates downward, possibly making them overly conservative. This type of limitation extends to the measure of general cognitive functioning. Because the general cognitive functioning measure was also administered orally and at times required verbal responses, this confound should be highly correlated with 24-month-old children's oral vocabularies. Consequently, use of a brief survey of parental report of 50 words spoken by their 24-month-old children, as well as statistical control for correlated verbal intelligence, may have resulted in very conservative estimates of oral vocabulary's "true" rela-

Table 4

Parameter Estimates (Standardized) of OLS Regressions Predicting Children's Behavioral Functioning at 60 months, ECLS–B Data

	Approaches to learning			Internalizing problems			Externalizing problems		
Variables	Model 1[a]	Model 2[b]	Model 3[c]	Model 1[a]	Model 2[b]	Model 3[c]	Model 1[a]	Model 2[b]	Model 3[c]
Intercept	0.02	0.34***	0.30***	–0.03	–0.17*	–0.15	0.03	–0.42***	–0.40***
Oral vocabulary									
24 months	0.22***	0.15***	0.10***	–0.11***	–0.10***	–0.09**	–0.14***	–0.07***	–0.06*
Race									
Black		0.05	0.08		–0.15*	–0.16*		0.04	0.03
Hispanic		0.04	0.06		–0.04	–0.05		–0.13*	–0.14*
Asian		–0.06	–0.05		0.03	0.02		–0.16*	–0.16*
Native American		–0.02	0.03		0.06	0.04		0.32	0.31
Other race		–0.002	0.02		0.03	0.03		–0.03	–0.03
Child age									
24 month		–0.03	–0.03		0.05*	0.06*		0.003	0.01
Gender									
Male		–0.22***	–0.21***		0.02	0.02		0.42***	0.42***
SES quintile									
Lowest		–0.16	–0.15		0.19*	0.19*		0.11	0.11
Second lowest		–0.20*	–0.18*		0.22**	0.22**		0.14	0.14
Middle		–0.12*	–0.11		0.26***	0.25***		0.06	0.05
Second highest		–0.05	–0.04		0.10	0.10		0.04	0.03
Nonsingleton		0.05	0.06		–0.04	–0.04		–0.05	–0.05
Mother's age at child's birth									
Equal to or older than 35		–0.02	–0.03		0.01	0.01		–0.01	–0.001
Younger than or equal to 18		–0.08	–0.09		–0.03	–0.03		0.04	0.04
Marital status									
Not married		–0.21**	–0.21**		–0.05	–0.05		0.23***	0.23**
Primary language									
Non-English		0.02	0.04		–0.14	–0.15		–0.11	–0.12
Birth weight									
Very low		–0.24***	–0.19**		–0.01	–0.03		–0.06	–0.07
Moderately low		–0.06	–0.04		0.001	–0.005		–0.03	–0.04
Labor complications		–0.01	–0.01		0.11**	0.11**		0.02	0.02
Obstetric procedures		–0.04	–0.04		0.001	–0.0001		–0.02	–0.02
Risks									
Medical		–0.08	–0.08		0.03	0.03		0.13*	0.12*
Behavioral		–0.13	–0.12		–0.06	–0.07		0.18	0.17
Parenting score									
24 months		0.08**	0.06*		–0.01	–0.005		–0.04	–0.03
Family member status									
Mental illness		–0.15	–0.14		0.04	0.04		0.13	0.13
Learning disability		–0.02	–0.01		0.11	0.11		0.11	0.11
Maternal health problems		–0.13	–0.13		0.22*	0.22*		–0.03	–0.03
Household status									
Special need		–0.06	–0.05		0.02	0.02		0.06	0.06
Mother depressed		–0.11	–0.11		0.02	0.02		0.15*	0.15*
Stayed in day-care center > 10 hr per week		–0.14*	–0.15*		–0.06	–0.06		0.30***	0.31***
Child television usage									
Middle third		0.10*	0.10*		–0.03	–0.03		–0.004	–0.003
Highest third		0.09	0.10*		–0.10	–0.11		–0.05	–0.05
Mother isolated		–0.04*	–0.04*		0.01	0.01		0.04	0.04
Approaches to learning									
24 months		0.12*	0.06						

Table 4
Continued

	Approaches to learning			Internalizing problems			Externalizing problems		
Variables	Model 1[a]	Model 2[b]	Model 3[c]	Model 1[a]	Model 2[b]	Model 3[c]	Model 1[a]	Model 2[b]	Model 3[c]
Internalizing problems									
24 months					0.15*	0.14			
Externalizing problems									
24 months								0.27**	0.25**
General cognitive functioning									
24 months			0.11**			−0.03			−0.04

Note. $N = 4{,}350$. Sample size rounded to nearest 50 per ECLS–B confidentiality requirements. OLS = ordinary least squares; ECLS–B = Early Childhood Longitudinal Study–Birth Cohort; SES = socioeconomic status.

[a] Intercept and oral vocabulary as predictors. [b] Adds race, age, gender, SES, nonsingleton, maternal age, language spoken in the home, birth weight, labor complications, obstetric procedures, medical and behavioral risks, parenting, family and maternal and household status, maternal depression, child care, television usage, maternal isolation, behavioral self-regulation and problem behaviors as predictors. [c] Adds general cognitive functioning as a predictor.

$*p < .05$. $**p < .01$. $***p < .001$.

tion with later academic and behavioral functioning. Second, and again due to the ECLS–B's design and measurement limitations, we were unable to directly contrast oral vocabulary's predictive utility against a variety of other early language or numeracy competencies (e.g., phonological awareness, number sense, working memory) that may also be predictive of later academic and behavioral functioning (e.g., Fitzpatrick & Pagani, 2012). However, many of these competencies should be correlated with other controls included in our study (e.g., SES, prior behavioral and general cognitive functioning). Third, a measure of expressive but not receptive vocabulary was available in the ECLS–B's 24-month survey wave. Therefore, we were unable to directly contrast the relative contribution of expressive versus receptive vocabulary knowledge to children's academic or behavioral functioning at school entry. However, strong positive correlations have been found between receptive and expressive vocabulary (e.g., $r = .66$; Sideridis & Simos, 2010), and Tomblin and Zhang's (2006) large-scale investigation of young children's language abilities yielded no evidence for a receptive–expressive vocabulary dissociation. Measuring children's ability to speak words should also better index the relative size of their vocabularies than measuring their ability to comprehend spoken words (Sénéchal et al., 2006). Fourth, although other studies indicate that academic and behavioral functioning by school entry contribute to long-term educational and societal opportunities (e.g., Chetty et al., 2011), we were unable to directly evaluate whether oral vocabulary continued to contribute to children's academic and behavioral functioning as they progressed through elementary school. This is because the ECLS–B's data collection ended at kindergarten entry. Fifth, the ECLS–B's data collection did not allow us to report on the quality and quantity of children's overall language exposure in the home. Language input in the home may also explain the relation between family SES and children's oral vocabulary (Fernald et al., 2013; Hart & Risley, 1995).

Study's Contributions and Implications

Oral vocabulary is theorized to contribute to many indicators of children's academic and behavioral functioning, including reading and mathematics achievement, behavioral self-regulation, and frequency of externalizing and internalizing problem behaviors (e.g., LeFevre et al., 2010; Perfetti & Stafura, 2014; Qi & Kaiser, 2004). Meaningful differences in children's oral vocabularies have been reported to occur by 24 months of age (e.g., Fernald et al., 2013). These differences have been hypothesized to strongly contribute to later achievement gaps (Hart & Risley, 1995; Hoff, 2013). Yet methodological and substantive limitations in existing work have been identified (NELP, 2008), and oral vocabulary has sometimes been reported to fail to uniquely predict children's achievement (Schatschneider et al., 2004). The result is substantial ambiguity as to oral vocabulary's theoretical and practical importance as a contributor to children's academic and behavioral functioning, and whether early interventions—as is increasingly suggested (e.g., Dickinson et al., 2010; Lesaux, 2012)—should emphasize preventing or remediating early vocabulary gaps to accelerate at-risk children's developmental trajectories.

Our study therefore has several theoretical and practical implications. We identify by the 24-month time period a wide range of sociodemographic, gestational and birth, family risk and resilience, and child-level factors associated with larger or smaller oral vocabularies that themselves may be included in early screening, monitoring, and intervention efforts. Consistent with prior research (e.g., Fernald et al., 2013; Hart & Risley, 1995), we initially found that family SES initially has a strong, consistent association with 24-month-old children's oral vocabularies. However, and which has been not previously reported, we subsequently found that this association is partially explained by children's own level of behavioral and general cognitive functioning in addition to other factors. Low SES in particular is increasingly recognized as being associated with fewer oral vocabulary-building opportunities (Mani, Mullainathan, Shafir, & Zhao, 2013) and child-directed speech (Weisleder & Fernald, 2013). Our study's findings of low SES's continuing negative relation with oral vocabulary, as well as the uniquely positive relation observed for parenting behaviors that are cognitively stimulating, child centered, and positive, are consistent with these theoretical accounts and suggest that interventions may need to be specifically targeted to 24-month-old children being raised in disadvantaged home environments (Hart & Risley, 1995; Rowe, 2008). Our findings are also consistent with other research indicating that being born with very low birth weight (Stolt, Haataja, Lapinleimu, & Lehtonen, 2008) and being raised in a household with a mentally ill family member (Zajicek-Farber, 2010) may contribute to delays in oral vocabulary acquisition. The negative association with being a nonsingleton is consistent with other work (Hillemeier, Farkas, Morgan, Martin, & Maczuga, 2009) and may result from parents speaking fewer words to each child when they are raising two or more children of the same age. Our findings indicate that the negative association with mental illness may not be unique to maternal depression, which itself is explained by children's own lower behavioral and general cognitive functioning, but may instead be related to a more general set of health and well-being indicators including a family member having a mental illness, being raised by a mother experiencing health problems, and being raised by a mother who feels socially isolated.

Our findings are consistent with prior work hypothesizing that a larger oral vocabulary helps children experience greater achievement and behavior (e.g., Perfetti & Stafura, 2014), thereby increasing the children's educational and societal opportunities as they age (e.g., Hoff, 2013; Lesaux, 2012). However, and across the study's five indicators of academic and behavioral functioning, other factors were also predictive. Consistent with this other work, these other factors included SES, low birth weight, watching television frequently, frequently being in child care (NICHD Early Child Care Research Network, 2003), and children's own prior level of general cognitive functioning. Our study's results extend prior work by suggesting that attending child care may also positively impact children's oral vocabularies. Our findings are consistent with prior work suggesting that parents who are stressed, overburdened, less engaged, and experiencing less social support may less frequently talk, read, or otherwise interact with their children (e.g., Paulson, Keefe, & Leiferman, 2009), resulting in these children having smaller oral vocabularies.

Prior empirical studies have largely been limited to investigations of single developmental pathways between vocabulary and (a) reading achievement, (b) mathematics achievement, or (c) behavioral functioning. Our study extends the field's knowledge base by rigorously establishing that oral vocabulary uniquely predicts multiple developmental pathways simultaneously and that this is evident as children are making the major transition to formal schooling. That oral vocabulary displays both temporal precedence and continued predictive utility despite extensive covariate adjustment across such a diverse number of indicators of children's development functions as a type of internal replication, providing empirical support to oral vocabulary's theorized importance as a first-order factor that may contribute to their early life-course trajectories. Our study also provides additional empirical support for claims that preventing or reducing later academic achievement gaps and their sequela may necessitate special emphasis on increasing at-risk children's oral vocabularies. Such efforts may need to occur before children are 24 months old. This is because oral vocabulary gaps are evident even by this very early time period, and in turn consistently predict children's academic and behavioral functioning as they begin kindergarten in the United States.

References

Andreassen, C., Fletcher, P., & Park, J. (2007). *Early Childhood Longitudinal Study, Birth Cohort (ECLS-B): Psychometric report for the 2-year data collection*. Washington, DC: National Center for Education Statistics.

Anthopolos, R., Edwards, S. E., & Miranda, M. L. (2013). Effects of maternal prenatal smoking and birth outcomes extending into the normal range on academic performance in fourth grade in North Carolina, USA. *Pediatric and Perinatal Epidemiology*, *27*, 564–574. doi:10.1111/ppe.12081

Bayley, N. (1993). *Bayley Scales of Infant Development* (2nd ed.). New York, NY: Psychological Corporation.

Bradley, R. H., & Caldwell, B. M. (1984). The HOME inventory and family demographics. *Developmental Psychology*, *20*, 315–320. doi:10.1037/0012-1649.20.2.315

Breaux, R. P., Harvey, E. A., & Lugo-Candelas, C. I. (2013). The role of parent psychopathology in the development of preschool children with behavior problems. *Journal of Clinical Child and Adolescent Psychology*, *43*, 777–790. doi:10.1080/15374416.2013.836451

Buck, K. R. (1997). *A comparison of three measures of social/emotional development of infants, toddlers, and preschoolers*. Unpublished dissertation, University of Alabama, Birmingham.

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Whitmore Schanzenbach, D., & Yagan, D. (2011). How does your kindergarten classroom affect your earnings? Evidence from Project STAR. *Quarterly Journal of Economics*, *126*, 1593–1660. doi:10.1093/qje/qjr041

Christakis, D. A., Zimmerman, F. J., DiGiuseppe, D. L., & McCarty, C. A. (2004). Early television exposure and subsequent attentional problems in children. *Pediatrics*, *113*, 708–713. doi:10.1542/peds.113.4.708

Cole, P. M., Armstrong, L. M., & Pemberton, C. K. (2010). The role of language in the development of emotion regulation. In S. D. Calkins & M. A. Bell (Eds.), *Child development at the intersection of emotion and cognition* (pp. 59–77). Washington, DC: American Psychological Association.

Davidse, N. J., De Jong, M. T., & Bus, A. G. (2014). Explaining common variance shared by early numeracy and literacy. *Reading and Writing*, *27*, 631–648. doi:10.1007/s11145-013-9465-0

DeNavas-Walt, C., Proctor, B. D., & Smith, J. C. (2013). *Income, poverty, and health insurance coverage in the United States: 2012* (U.S. Census Bureau No. P60–245). Washington, DC: US Government Printing.

Dickinson, D. K., Golinkoff, R. M., & Hirsh-Pasek, K. (2010). Speaking out for language: Why language is central to reading development. *Educational Researcher*, *39*, 305–310. doi:10.3102/0013189X10370204

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, *43*, 1428–1446. doi:10.1037/0012-1649.43.6.1428

Eisenberg, N., Sadovsky, A., & Spinrad, T. L. (2005). Associations of emotion-related regulation with language skills, emotion knowledge, and academic outcomes. *New Directions for Child and Adolescent Development*, *109*, 109–118. doi:10.1002/cd.143

Ennemoser, M., & Schneider, W. (2007). Relations of television viewing and reading: Findings from a 4-year longitudinal study. *Journal of Educational Psychology*, *99*, 349–368. doi:10.1037/0022-0663.99.2.349

Farkas, G., & Beron, K. (2004). The detailed age trajectory of oral vocabulary knowledge: Differences by class and race. *Social Science Research*, *33*, 464–497. doi:10.1016/j.ssresearch.2003.08.001

Fenson, L., Dale, P., Reznick, J., Bates, E., Thal, D., & Pethick, S. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, *59*(5). doi:10.2307/1166093

Fenson, L., Dale, P., Reznick, J., Thal, D., & Bates, E. (1993). *The Macarthur Communicative Development Inventories: User's guide and technical manual*. Baltimore, MD: Brookes.

Fernald, A., Marchman, V. A., & Weisleder, A. (2013). SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental Science*, *16*, 234–248. doi:10.1111/desc.12019

Fitzpatrick, C., & Pagani, L. S. (2012). Toddler working memory skills predict kindergarten school readiness. *Intelligence*, *40*, 205–212. doi:10.1016/j.intell.2011.11.007

Ganger, J., & Brent, M. R. (2004). Reexamining the vocabulary spurt. *Developmental Psychology*, *40*, 621–632. doi:10.1037/0012-1649.40.4.621

Guo, G., & Harris, K. M. (2000). The mechanisms mediating the effects of poverty on children's intellectual development. *Demography*, *37*, 431–447. doi:10.1353/dem.2000.0005

Hack, M., Klein, N. K., & Taylor, H. G. (1995). Long-term developmental outcomes of low birth weight infants. *Future of Children*, *5*, 176–196. doi:10.2307/1602514

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Brookes.

Heckman, J. J., & Masterov, D. V. (2007). The productivity argument for investing in young children. *Review of Agricultural Economics*, *29*, 446–493. doi:10.1111/j.1467-9353.2007.00359.x

Hillemeier, M., Farkas, G., Morgan, P. L., Martin, M., & Maczuga, S. (2009). Disparities in the prevalence of cognitive delay: How early do they appear? *Pediatric & Perinatal Epidemiology*, *23*, 186–198. doi:10.1111/j.1365-3016.2008.01006.x

Hoff, E. (2013). Interpreting the early language trajectories of children from low-SES and language minority homes: Implications for closing achievement gaps. *Developmental Psychology*, *49*, 4–14. doi:10.1037/a0027238

Hoff, E., Laursen, B., & Bridges, K. (2012). Measurement and model building in studying the influence of socioeconomic status on child development. In M. Lewis & L. Mayes (Eds.), *A developmental environmental measurement handbook* (pp. 590–606). New York, NY: Cambridge University Press.

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing*, *2*, 127–160. doi:10.1007/BF00401799

Horwitz, S. M., Irwin, J. R., Briggs-Gowan, M. J., Bosson Heenan, J. M., Mendoza, J., & Carter, A. S. (2003).

Language delay in a community cohort of young children. *Journal of the American Academy of Child and Adolescent Psychiatry, 42*, 932–940. doi:10.1097/01.CHI.0000046889.27264.5E

Iruka, I. U., LaForett, D. R., & Odom, E. C. (2012). Examining the validity of the family investment and stress models and relationship to children's school readiness across five cultural groups. *Journal of Family Psychology, 26*, 359–370. doi:10.1037/a0028290

Klebanov, P. K., Brooks-Gunn, J., & McCormick, M. C. (1994). Classroom behavior of very low birth weight elementary school children. *Pediatrics, 94*, 700–708.

Koolstra, C., & Van der Voort, T. (1996). Longitudinal effects of television on children's leisure time reading: A test of three explanatory models. *Human Communication Research, 23*, 4–35. doi:10.1111/j.1468-2958.1996.tb00385.x

LeFevre, J., Fast, L., Skwarchuk, S., Smith-Chant, B., Bisanz, J., Kamawar, D., & Penner-Wilger, M. (2010). Pathways to mathematics: Longitudinal predictors of performance. *Child Development, 81*, 1753–1767. doi:10.1111/j.1467- 8624.2010.01508.x

Lesaux, N. K. (2012). Reading and reading instruction for children from low-income and non-English-speaking households. *The Future of Children, 22*, 73–88. doi:10.1353/foc.2012.0010

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*, 1–75. doi:10.1017/S0140525X99001776

Li-Grining, C., Votruba-Drzal, E., Maldonado-Carreno, C., & Haas, K. (2010). Children's early approaches to learning and academic trajectories through fifth grade. *Developmental Psychology, 46*, 1062–1077. doi:10.1037/a0020066

Lynch, J. L. (2011). Infant health, race/ethnicity, and early educational outcomes using the ECLS-B. *Sociological Inquiry, 81*, 499–526. doi:10.1111/j.1475-682X.2011.00390.x

Mani, A., Mullainathan, S., Shafir, E., & Zhao, J. (2013). Poverty impedes cognitive function. *Science, 341*, 976–980. doi:10.1126/science.1238041

McLoyd, V. C. (1998). Socioeconomic disadvantage and child development. *American Psychologist, 53*, 185–204. doi:10.1037/0003-066X.53.2.185

McMurray, B. (2007). Defusing the vocabulary explosion. *Science, 317*, 631. doi:10.1126/science.1144073

Menting, B., Van Lier, P. A. C., & Koot, H. M. (2010). Language skills, peer rejection, and the development of externalizing behavior from kindergarten to fourth grade. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 52*, 72–79. doi:10.1111/j.1469-7610.2010.02279

Mitchell, A. M., & Brady, S. A. (2013). The effect of vocabulary knowledge on novel word identificaiton. *Annals of Dyslexia, 63*, 201–216. doi:10.1007/s11881-013-0080-1

Morgan, P. L., Farkas, G., & Wu, Q. (2009). Five-year growth trajectories of kindergarten children with learning difficulties in mathematics. *Journal of Learning Disabilities, 42*, 306–321. doi:10.1177/0022219408331037

Moyé, L. E. (2008). The multiple comparison issue in health care research. In C. R. Rao, J. P. Miller, & D. C. Rao (Eds.), *Handbook of statistics: Epidemiology and medical statistics* (Vol. 27, pp. 616–655). New York, NY: Elsevier.

Najarian, M., Snow, K., Lennon, J., & Kinsey, S. (2010). *Early Childhood Longitudinal Study, Birth Cohort (ECLS-B), preschool–kindergarten 2007 psychometric report* (NCES 2010-009). Washington, DC: National Center for Education Statistics.

National Early Literacy Panel. (2008). *Developing early literacy: Report of the National Early Literacy Panel*. Jessup, MD: National Institute for Literacy.

National Institute of Child Health and Human Development Early Child Care Research Network. (2003). Does amount of time spent in childcare predict socioemotional adjustment during the transition to kindergarten? *Child Development, 74*, 976–1005.

National Institute of Child Health and Human Development Early Child Care Research Network. (2005). Pathways to reading: The role of oral language in the transition to reading. *Developmental Psychology, 41*, 428–442. doi:10.1037/0012-1649.41.2.428

Nord, C., Edwards, B., Andreassen, C., Green, J., & Wallner-Allen, K. (2006). *Early Childhood Longitudinal Study, Birth Cohort (ECLS-B), User's manual for the ECLS-B longitudinal 9-month-2-year data file and electronic codebook* (NCES 2006–046). Washington, DC: National Center for Education Statistics.

Ouellette, G. P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology, 98*, 554–566. doi:10.1037/0022-0663.98.3.554

Pagani, L. S., Fitzpatrick, C., & Barnett, T. A. (2013). Early childhood television viewing and kindergarten entry readiness. *Pediatric Research, 74*, 350–355. doi:10.1038/pr.2013.105

Pan, B. A., Rowe, M. L., Singer, J. D., & Snow, C. E. (2005). Maternal correlates of growth in toddler vocabulary production in low-income families. *Child Development, 76*, 763–782. doi:10.1111/1467-8624.00498-i1

Paulson, J. F., Keefe, H. A., & Leiferman, J. A. (2009). Early parental depression and child language development. *Journal of Child Psychology and Psychiatry, 50*, 254–262. doi:10.1111/j.1469-7610.2008.01973.x

Pennington, B. F., & Bishop, D. V. M. (2009). Relations among speech, language, and reading disorders. *Annual Review of Psychology, 60*, 283–306. doi:10.1146/annurev.psych.60.110707.163548

Perfetti, C. A., & Hart, L. (2001). The lexical basis of comprehension skill. In D. S. Gorfien (Ed.), *On the consequences of meaning selection: Perspectives on resolving lexical ambiguity* (pp. 67–86) Washington, DC: American Psychological Association.

Perfetti, C., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading, 18*, 22–37. doi:10.1080/10888438.2013.827687

Purpura, D. J., Hume, L. E., Sims, D. M., & Lonigan, C. J. (2011). Early literacy and early numeracy: The value of including early literacy skills in the prediction of numeracy development. *Journal of Experimental Child Psychology, 110*, 647–658. doi:10.1016/j.jecp.2011.07.004

Qi, C. H., & Kaiser, A. P. (2004). Problem behaviors of low-income children with language delays: An observation study. *Journal of Speech, Language, and Hearing Research, 47*, 595–610. doi:10.1044/1092-4388(2004/046)

Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1*, 385–401. doi:10.1177/014662167700100306

Raghunathan, T. E., Solenberger, P. W., & Van Hoewyk, J. (2002). *IVEware: Imputation and variance estimation software*. Ann Arbor, MI: Survey Methodology Program, Survey Research Center, Institute for Social Research, University of Michigan.

Rescorla, L., Ross, G. S., & McClure, S. (2007). Language delay and behavioral/emotional problems in toddlers: Findings from two developmental clinics. *Journal of Speech, Language, and Hearing Research, 50*, 1063–1078. doi:10.1044/1092-4388(2007/074)

Rowe, M. L. (2008). Child-directed speech: Relation to socioeconomic status, knowledge of child development and child vocabulary skill. *Journal of Child Language, 35*, 185–205. doi:10.1017/S0305000907008343

Sabol, T. J., & Pianta, R. C. (2012). Patterns of school readiness forecast achievement and socioemotional development at the end of elementary school. *Child Development, 83*, 282–299. doi:10.1111/j.1467-8624.2011.01678.x

Schatschneider, C., Fletcher, J., Francis, D., Carlson, C., & Foorman, B. (2004). Kindergarten prediction of reading skills: A longitudinal comparative analysis. *Journal of Educational Psychology, 96*, 265–282. doi:10.1037/0022-0663.96.2.265

Sénéchal, M., Ouellette, G., & Rodney, D. (2006). The misunderstood giant: On the predictive role of early vocabulary to future reading. In S. B. Neuman, & D. Dickinson (Eds.), *Handbook of early literacy research* (Vol. 2, pp. 173–182). New York, NY: Guilford.

Sideridis, G. D., & Simos, P. (2010). What is the actual correlation between expressive and receptive measures of vocabulary? Approximating the sampling distribution of the correlation coefficient using the bootstrapping method. *International Journal of Educational and Psychological Assessment, 5*, 117–133.

Silven, M., Poskiparta, E., Niemi, P., & Voeten, M. (2007). Precursors of reading skill from infancy to first grade in Finnish: Continuity and change in a highly inflected language. *Journal of Educational Psychology, 99*, 516–531. doi:10.1037/0022-0663.99.3.516

Skarakis-Doyle, E., Campbell, W., & Dempsey, L. (2009). Identification of children with language impairment: Investigating the classification accuracy of the MacArthur-Bates Communicative Development Inventories, Level III. *American Journal of Speech Language Pathology, 18*, 277–288. doi:10.1044/1058-0360(2009/08- 0035)

Spelke, E. S., & Tsivkin, S. (2001). Language and number: A bilingual training study. *Cognition, 78*, 45–88. doi:10.1016/S0010-0277(00)00108-6

Stolt, S., Haataja, L., Lapinleimu, H., & Lehtonen, L. (2008). The early lexical development and its predictive value to language skills at 2 years in very-low-birth-weight children. *Journal of Communication Disorders, 42*, 107–123. doi:10.1016/j.jcomdis.2008

Storch, S. A., & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology, 38*, 934–947. doi:10.1037/0012-1649.38.6.934

Tomblin, B. J., & Zhang, X. (2006). The dimensionality of language ability in school-age children. *Journal of Speech, Language, and Hearing Research, 49*, 1193–1208. doi:10.1044/1092-4388(2006/086)

Vaughn, B. E., Kopp, C. B., & Krakow, J. B. (1984). The emergence and consolidation of self-control from eighteen to thirty months of age: Normative trends and individual differences. *Child Development, 55*, 990–1004. doi:10.2307/1130151

Verhoeven, L., vanLeeuwe, J. &., & Vermeer, A. (2011). Vocabulary growth and reading development across the elementary school years. *Scientific Studies of Reading, 15*, 8–25. doi: 10.1080/10888438.2011.536125

Walker, D., Greenwood, C., Hart, B., & Carta, J. (1994). Prediction of school outcomes based on early langauge production and socioeconomic factors. *Child Development, 65*, 606–621. doi:10.2307/1131404

Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science, 24*, 2143–2152. doi:10.1177/0956797613488145

Winsler, A., Diaz, R. M., McCarthy, E. M., Atencio, D. J., & Chabay, L. A. (1999). Mother-child interaction, private speech, and task performance in preschool children with behavior problems. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 40*, 891–904. doi:10.1111/1469-7610.00507

Zajicek-Farber, M. L. (2010). The contributions of parenting and postnatal depression on emergent language of children in low-income families. *Journal of Child and Family Studies, 19*, 257–269. doi:10.1007/s10826-009-9293-7

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's website:

Appendix S1. Note on Oral Vocabulary Terminology

Appendix S2. Additional Knowledge Base Limitations

Appendix S3. Multiplicity Check Information

Child Development, September/October 2015, Volume 86, Number 5, Pages 1371–1385

Flexibility in Bilingual Infants' Word Learning

Katharine Graf Estes
University of California, Davis

Jessica F. Hay
University of Tennessee, Knoxville

The present experiments tested bilingual infants' developmental narrowing for the interpretation of sounds that form words. These studies addressed how language specialization proceeds when the environment provides varied and divergent input. Experiment 1 ($N = 32$) demonstrated that bilingual 14- and 19-month-olds learned a pair of object labels consisting of the same syllable produced with distinct pitch contours (rising and falling). Infants' native languages did not use pitch contour to differentiate words. In Experiment 2 ($N = 16$), 22-month-old bilinguals failed to learn the labels. These results conflict with the developmental trajectory of monolinguals, who fail to learn pitch contour contrasts as labels at 17–19 months (Hay, Graf Estes, Wang, & Saffran, 2015). Bilingual infants exhibited a prolonged period of flexibility in their interpretation of potential word forms.

A majority of children worldwide learn more than one language (Grosjean, 2010). Yet, much of what we understand about early language acquisition is derived from studies of monolingual children. There are many broad similarities between monolingual and bilingual development that highlight the impressive nature of bilingual learning. For example, despite the additional computational demands of acquiring two separate linguistic systems, bilinguals meet many language milestones, such as first words, at around the same ages as monolinguals (e.g., Oller, Eilers, Urbano, & Cobo-Lewis, 1997; Pearson, Fernandez, & Oller, 1993; Petitto et al., 2001). Furthermore, monolinguals' and bilinguals' total vocabulary sizes are comparable when all of the words bilinguals know across languages are included in the count (Hoff et al., 2012; Pearson et al., 1993; Petitto et al., 2001). Beyond these gross similarities, young bilinguals and monolinguals display some intriguing differences in their patterns of development (see reviews by Byers-Heinlein & Fennell, 2014; Sebastián-Gallés, 2011). In the present experiments, we explore the possibility that bilingual experience may promote extended flexibility in early word learning.

At the phonological level, monolingual and bilingual infants must discover the phoneme boundaries, phoneme combinations, allophonic variations, word boundaries, and the rhythmic patterns of the words and sentences of their native language(s). This is a demanding process because infants have no a priori knowledge of which sound variants (including phonemes and allophones, as well as pitch, duration, and phonotactic patterns) are relevant in their native languages, or how these sound variants will be used. Bilingual infants, however, have the added challenge of determining how sound variants are used in two systems instead of one. In the bilingual literature, there are inconsistent findings regarding the time course of phonological development, and the course of perceptual narrowing, more specifically. In language acquisition, *perceptual narrowing* entails two complementary processes: perceptual attenuation of some non-native phoneme contrasts coupled with perceptual maintenance or enhancement of native-language contrasts (e.g., Kuhl et al., 2006; Werker & Tees, 1984). There are findings that suggest bilinguals take longer than monolinguals to home in on language-specific phoneme categories (Bosch & Sebastián-Gallés, 2003b; Sebastián-Gallés & Bosch, 2009). For example, across the 1st year of life, bilingual Spanish-Catalan infants show a U-shaped function

---

This research was supported by a grant to Katharine Graf Estes from the National Science Foundation (BCS0847379). We would like to thank Carolina Bastos, Stephanie Chen-Wu Gluck, Dylan Antovich, and the members of the Language Learning Lab at the University of California, Davis, for their assistance with this research. Thanks also to Lisa Oakes and Casey Lew-Williams for helpful discussion and comments. We also thank the parents who generously contributed their time.

Correspondence concerning this article should be addressed to Katharine Graf Estes, Department of Psychology, University of California, Davis, 1 Shields Ave., Davis, CA 95616. Electronic mail may be sent to kgrafestes@ucdavis.edu.

© 2015 The Authors
Child Development © 2015 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2015/8605-0005
DOI: 10.1111/cdev.12392

for both Catalan vowel (i.e., /e/–/ɛ/contrast) discrimination (Bosch & Sebastián-Gallés, 2003b) and consonant (/s/–/z/contrast) discrimination (Bosch & Sebastián-Gallés, 2003a). At an intermediate age, bilingual infants appear to lose the ability to discriminate contrasts that younger and older infants can discriminate. Monolinguals do not show this same U-shaped function. Bosch and Sebastián-Gallés (2003b) proposed that monolingual and bilingual input produce different courses for the language-specific reorganization of the perceptual system. As discussed in more detail below, bilinguals receive less input in a given language than monolinguals receive, and the phoneme category boundaries within and across languages are less clear than in monolingual input. Bilinguals may need to accumulate additional exposure to form strong expectations about phoneme categories. However, other evidence indicates that bilinguals and monolinguals show similar patterns of development in speech perception (Albareda-Castellot, Pons, & Sebastián-Gallés, 2011; Burns, Yoshida, Hill, & Werker, 2007; Sundara, Polka, & Molnar, 2008). The observed patterns seem to depend greatly on the procedures and measures implemented, the attentional demands required to solve the task, and the languages and types of contrasts studied (for reviews, see Byers-Heinlein & Fennell, 2014; Curtin, Byers-Heinlein, & Werker, 2011; Sebastián-Gallés, 2011).

The varied behavioral findings regarding phonological development leave open the possibility that perceptual narrowing takes place over different developmental periods for bilingual and monolingual infants. There is recent neural evidence supporting this possibility. In an experiment using event-related brain potentials, Garcia-Sierra et al. (2011) tested Spanish-English bilingual infants' discrimination of Spanish and English consonants. Bilingual infants exhibited neural discrimination at 10–12 months of age, but not at 6–9 months of age. In contrast, English-learning monolinguals showed neural discrimination of both native and non-native contrasts at 7 months of age, but only of native contrasts at 11 months (Rivera-Gaxiola, Silva-Pereyra, & Kuhl, 2005). Garcia-Sierra et al. (2011) suggested that it may take longer for bilingual infants to establish language-specific neural commitment from their broad (language-universal) perceptual abilities. Further support for this notion was reported in an experiment using functional near infrared spectroscopy. Petitto et al. (2012) indicated that across development, bilingual infants tended to show similar patterns of activation of left inferior frontal cortex for native and non-native (Hindi) phoneme contrasts. In contrast, across the same developmental period (around 4–6 months vs. 10–12 months), monolinguals' activation diverged for native and non-native contrasts.

There are several reasons why bilingual phonological development may show different patterns than monolingual development (Byers-Heinlein & Fennell, 2014; Costa & Sebastián-Gallés, 2014; Curtin et al., 2011). First, bilingual infants likely hear less speech in each language than monolinguals hear in a single language. Second, there are greater computational demands in learning two separate phonological, lexical, and syntactic systems than in learning a single system. Infants must track information in each language separately to learn the distinct systems. They must also discern, moment to moment, which language is relevant in the immediate context. Third, the languages may divide the same acoustic space into different categories; bilinguals may be required to interpret conflicting cues to sound structure or interpret the same sound differently depending on the context. Finally, Byers-Heinlein and Fennell (2014) described bilingual input as "noisy" and highly variable (see also Garcia-Sierra et al., 2011). One source of noise occurs because many bilingual infants hear two different languages produced by the same person, sometimes even within the same sentence (Byers-Heinlein, 2013), thereby potentially making it difficult to process sound patterns separately in each language. In addition, many bilingual children learn from parents who are themselves bilingual. Thus, there may be greater variability in the acoustic realization of some speech sounds in bilingual input (Sundara, Polka, & Baum, 2006) and an increase in the frequency of production errors compared to monolingual input (Bosch & Ramon-Casas, 2011). Noise in the input may make it challenging to determine how to weight acoustic-phonetic dimensions of the speech signal to focus on what is critical for differentiating between words with different meanings.

Infants' learning about the sound inventory of their language(s) ultimately serves the process of linking sounds (word forms) with meanings during word learning. How do bilingual infants apply their early learning about sounds to acquire words? Studies of early lexical development suggest that bilinguals differ in their attention to phonetic detail in words relative to monolinguals. For example, Ramon-Casas, Swingley, Sebastián-Gallés, and Bosch (2009) reported that during word recognition, bilingual toddlers did not detect mispronunciations of vowels found in highly familiar words when the vowels were contrastive in only one of their languages, whereas monolinguals readily detected such native mispronunciations in their own language.

Curtin et al.'s (2011) Processing Rich Information from Multidimensional Interactive Representations (PRIMIR) framework suggests that bilingual infants face different demands when mapping sounds to meanings than do monolinguals. Demands such as reduced input, crowded phonetic space, and language switching may tax the limited resources available for attending to fine phonetic detail while simultaneously mapping sounds to meaning. In a novel word learning task, Fennell, Byers-Heinlein, and Werker (2007) examined bilinguals' learning of minimal pair object labels that differed by a single phoneme (i.e., /bI/and/dI/). At 14 months of age, monolinguals typically fail to learn such labels in the absence of supportive referential context, but by 17 months they typically succeed even without the support (e.g., Werker, Fennell, Corcoran, & Stager, 2002). Fennell et al. (2007) found that it took bilingual infants an additional 3 months before they succeeded at the same task; 17-month-old bilinguals did not learn the labels, but by 20 months they succeeded (but see Fennell & Byers-Heinlein, 2014; Mattock, Polka, Rvachew, & Krehm, 2010, for task-related effects).

Similar to the development of perceptual narrowing, bilingual experience may affect word learning by requiring more time to develop strong representations of the ways that sounds can differentiate between words (i.e., abstract phoneme representations; Curtin et al., 2011; Werker & Curtin, 2005). Exposure to two phonological systems may require a protracted period of development to determine what information is relevant during word learning and what is not. That is, bilinguals may take more time than monolinguals to center on what sound variants are lexically contrastive and can therefore distinguish words. This may contribute to bilinguals' later learning of minimal pairs, relative to monolinguals in the equivalent task (Fennell et al., 2007; Werker et al., 2002). In addition, the developmental difference may mean that in some circumstances, bilinguals may be *more* inclusive than monolinguals regarding the types of sound variants that can map to meanings, particularly if those sounds are acoustically salient and have not been assimilated to native sound categories.

In the present experiments, we test the hypothesis that bilingual experience may promote an extended period of flexibility during early word learning. Across domains, narrowing that occurs throughout development is thought to reflect specialization for processing the infant's particular environment (Scott, Pascalis, & Nelson, 2007). As discussed earlier, infants display perceptual narrowing for phoneme categories, focusing on the sounds that occur in their native language(s). Later in acquisition, infants also exhibit *interpretive narrowing* in word learning (Hay et al., 2015), in which they constrain the range of sounds (e.g., beeps, communicative vocal sounds) and sound sequences (e.g., phonotactically unattested combinations) that can be mapped to meanings (Graf Estes, Edwards, & Saffran, 2011; Namy, 2001; Woodward & Hoyne, 1999). The openness or flexibility that has been proposed in bilingual phoneme category perception (Garcia-Sierra et al., 2011; Petitto et al., 2012) may extend to word learning as well. Thus, the developmental course of interpretive narrowing in word learning may differ for monolinguals and bilinguals.

Although protracted interpretive narrowing may occur hand in hand with protracted perceptual narrowing, there are additional factors that may apply specifically to flexibility in word learning. First, there is a greater diversity in the potential forms that words can take in bilingual environments as compared to monolingual environments. The phonemic and phonotactic inventories across two languages are broader than the inventories of a single language, and this breadth of word forms may encourage bilingual infants to entertain a wide range of potential word forms.

There are also factors outside of the phonological system that could affect flexibility. Although both monolingual and bilingual infants learn to map sound sequences onto objects in their environments, bilinguals frequently must accept more than one label for an individual concept. Accordingly, bilinguals do not apply mutual exclusivity (or disambiguation) as stringently as monolinguals (Bialystok, Barac, Blaye, & Poulin-Dubois, 2010; Byers-Heinlein & Werker, 2013). Finally, because bilinguals must switch between languages, they must simultaneously activate one phonological system and inhibit the other, while maintaining connections between each language and an underlying conceptual system. Research suggests that this demanding process may lead to enhanced cognitive flexibility in early child development (Kovacs & Mehler, 2009; Poulin-Dubois, Blaye, Coutya, & Bialystok, 2011) and in early and later adulthood (Bialystok, Craik, & Luk, 2008) for tasks that require attentional control, shifting attention, and inhibition. Here, we explore the possibility that bilingual experience may also promote flexibility regarding the sounds that can be mapped to meanings in early word learning.

There are several important criteria for the selection of stimuli to produce a strong test of flexibility in the sounds that bilinguals attend to during word learning. First, the sound variants involved should be acoustically salient, so that any lack of label learning cannot be attributed to a failure to perceive differences between labels or due to the challenges of attending to small phonetic details while simultaneously mapping sounds to objects. Second, the labels should form naturalistic possible words, so that infants' prior experience with word forms is not violated. Third, to produce a naturalistic test, the labels should consist of sound variants that are used to differentiate words in some languages, but not in the infants' native languages.

Given these criteria, pitch contour serves as an excellent means to examine flexibility in word learning in bilinguals. Around 60%–70% of the world's languages are tonal languages (Yip, 2002), meaning that speakers can use differences in pitch contour (i.e., lexical tone) to differentiate word meanings. For example, in Mandarin Chinese the syllable /ma/ produced with a high level contour (Tone 1) means "mother," but produced with a low dipping contour (Tone 3) means "horse." For English learners, and learners of other nontonal languages, pitch contour is a highly salient and familiar component of the speech stream, but it is used quite differently than in tonal languages. Variations in pitch can signal emotion (Murray & Arnott, 1993), grammatical structure (Gussenhoven, 2004), lexical stress (Fry, 1958), speaker identity, emphasis, and speaking register (e.g., infant- vs. adult-directed speech; Fernald, 1992). Thus, infants have experience with pitch variation, but none of this experience indicates that words can differ by pitch contour alone.

In tone perception tasks, there is evidence that perceptual narrowing occurs for some tone contrasts quite early in development. For example, Yeung, Chen, and Werker (2013) presented 4- and 9-month-old English-learning infants with two distinct pitch contours (high rising and mid level tones) that are used to distinguish word meanings in Cantonese. They found that 4-month-olds, but not 9-month-olds, differentiated the tones (see also Mattock, Molnar, Polka, & Burnham, 2008). However, other tone contrasts appear to be resistant to perceptual narrowing. For example, Mattock and Burnham (2006) demonstrated that while English-learning infants lose the ability to discriminate Thai rising and low tones between 6 and 9 months of age, they continue to discriminate Thai rising and falling tones throughout this period (see also So & Best, 2010, for additional evidence of contrast-dependent lexical tone discrimination). Perceptual narrowing appears to depend, at least in part, on how acoustically distinctive or confusable the tones are. The distinctiveness of rising versus falling tones, for example, makes them more resistant to perceptual narrowing than some other contrasts.

Although lexical tones vary in how acoustically distinctive they are, infants learning nontonal languages should come to ignore pitch contours when mapping labels to objects. There should be interpretive narrowing regarding pitch contours in label learning because the input indicates that differences in pitch contour do not differentiate words. Two recent word learning studies provide evidence of changing sensitivity to pitch contour across the 2nd year of life. Hay et al. (2015) investigated monolingual English-learning 14-, 17-, and 19-month-olds' ability to learn a pair of object labels that differed only in their pitch contours (i.e., the syllable *ku* produced with a rising vs. falling contour). Using a modified version of the Switch paradigm (Werker, Cohen, Lloyd, Casasola, & Stager, 1998), infants were habituated to two label-object pairings, then tested on trials in which the original pairings were maintained and on trials in which the original pairings were switched. Longer looking on the test trials in which the pairings were switched indicated that the infants learned the labels. Hay et al. found that only 14-month-olds readily detected the altered test trials, indicating that they learned identical syllables with distinct pitch contours as separate object labels. Seventeen- and 19-month-olds did not. Thus, English-learning novice word learners are open to accepting object labels that differ only in pitch contour, whereas infants just a few months older have narrowed their interpretation of what sounds are lexically contrastive.

Singh, Hui, Chan, and Golinkoff (2014) found somewhat later interpretive narrowing for pitch contour in both monolinguals and bilinguals. They tested three groups of 18- and 24-month-olds: monolingual English learners, Mandarin learners also acquiring English, and bilinguals acquiring two nontonal languages. Within a referential context (i.e., "Look here! It's a ____!"), infants were familiarized with two object labels that differed in both initial consonant and pitch contour (*leng* rising vs. *beng* falling). During testing, the labels changed in pitch contour (i.e., *leng* rising → *leng* falling). Mandarin-learning 18- and 24-month-olds detected the altered pitch contours. However, English-learning infants and bilingual nontonal language learners detected the altered pitch contours at 18 months, but not at 24 months. Thus, younger nontonal language learners attended to pitch contour, treating it as a lexically contrastive feature. By 24 months, only infants who had experience with tonal languages attended to the alteration in the words' pitch contours.

Across studies, there is a discrepancy regarding monolingual infants' attention to pitch contour during word learning around 1.5 years of age. Hay et al. (2015) found evidence of interpretive narrowing between 14 and 17–19 months of age, whereas Singh et al. (2014) found continued flexibility at 18 months. Differences in the demands of each task may have produced the conflicting findings (Curtin et al., 2011; Werker & Curtin, 2005) and may affect our understanding of bilingual word learning as well. Singh et al.'s task provided naturalistic referential context in the labeling phrases and through the inclusion of familiar objects. The supportive context may have promoted attention to detail in the input (Fennell & Waxman, 2010), thereby supporting (and perhaps overestimating) infants' attention to the non-native use of pitch contour in the new words. Consistent with this notion, May and Werker (2014) found that a task with referential cues supported infants' learning of labels containing non-native click contrasts, whereas a Switch task lacking referential cues did not.

Singh et al.'s (2014) results suggest later interpretive narrowing for monolinguals (between 18 and 24 months) and equivalent flexibility or openness in processing for monolingual and bilinguals. Conversely, relatively pared down tasks that lack referential cues, like the Switch paradigm used by Hay et al. (2015) and others (e.g., Stager & Werker, 1997), may be better able to reveal the early stages of experience-induced interpretive narrowing because they provide less attentional push. Indeed, the monolingual English-learning infants in Hay et al.'s study processed pitch in a way that actually appears to be more sophisticated (i.e., ignoring pitch contour differences when learning a nontonal language), displaying interpretive narrowing by 17 months. In this case, the task lacking referential support revealed aspects of word learning that were hidden by the more supportive context (Singh et al., 2014).

The comparisons across the research and Hay et al. (2015) and Singh et al. (2014) illustrate two important ideas. First, although progress has been made (Curtin et al., 2011; Fennell & Waxman, 2010; May & Werker, 2014), we do not fully understand how the dynamics of experimental tasks affect language processing. Second, the comparisons demonstrate the importance of examining development using multiple methodologies. Based on this literature, it is not clear whether bilingual infants also undergo early interpretive narrowing for pitch contour because they have only been tested in a referential task (Singh et al., 2014), which may have overestimated similarities between monolinguals' and bilinguals' interpretation of pitch contour during word learning.

As discussed above, there are many reasons to expect that bilinguals might display a protracted period of flexibility in word learning. In particular, they may show interpretive narrowing at a later age than monolinguals because their phonological input is more variable and is divided across two distinct systems, and they receive experience with a broader range of ways that sounds can differentiate words than monolinguals experience. In addition to these potential causal mechanisms for a protracted period of flexibility, it would be functionally adaptive for bilinguals to remain open to accepting a wider variety of sound distinctions as labels for novel objects (see also Garcia-Sierra et al., 2011; Petitto et al., 2012). Importantly, a prolonged period of interpretive narrowing may leave bilinguals open to accepting less attested, but phonologically legal, distinctions as they continue to gather information about how sounds are used in each language.

In the present experiments, we investigated the trajectory of bilinguals' processing of a salient, communicatively relevant aspect of the speech signal that is not used to differentiate words in either of the infants' native languages. Experiment 1 examined bilingual 14- and 19-month-olds' interpretation of pitch contour as a means to differentiate words. The experiment tested whether nontonal language bilinguals show the same developmental shift as English-learning monolinguals when learning object labels that contrast only in pitch contour (Hay et al., 2015). We used the stimuli and methodology presented by Hay et al. (2015) to test whether, in the absence of referential support, bilinguals show early interpretive narrowing in word learning, like monolinguals in the same task, or whether they show a protracted period of attention to pitch contour. Infants were presented with a pair of novel objects and novel labels that consisted of the same syllable, *ku* (/kʊ/), produced with a rising pitch contour in one object label and with a falling pitch contour in the second object label.

If bilingual infants are not yet fully committed to the sound variants that differentiate words in their native languages, they may remain open to learning labels that differ only in pitch contour for longer than their monolingual peers. In Experiment 1, we predicted that bilingual infants would learn labels that differed only in pitch contour at both 14 and 19 months of age. In contrast, monolingual infants experience a shift in their interpretation of pitch contour between 14 and 19 months of age (Hay et al., 2015).

Experiment 1

Method

Participants

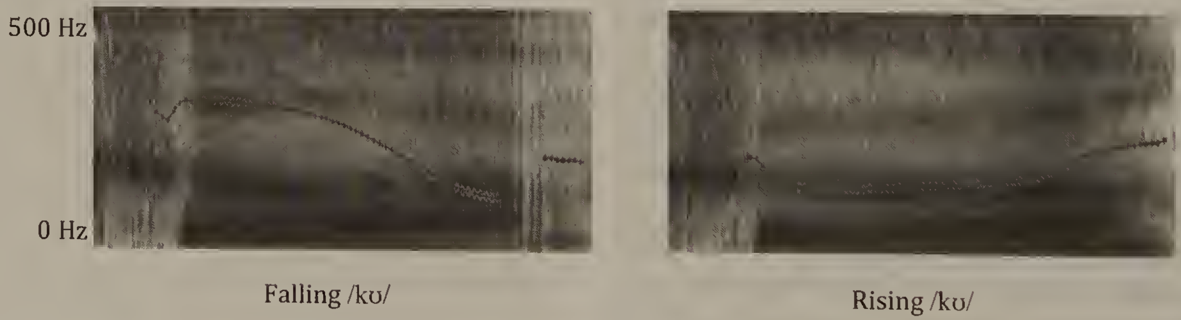
The participants were 16 bilingual infants aged 14 months (*M* = 14.5 months, range = 13.9–15.1 months; 8 female) and 16 bilingual infants aged 19 months (*M* = 19.7 months, range = 19.0–20.6; 8 female). All infants were born full term and had no history of hearing or vision impairments. Based on a parental report questionnaire and interview, each infant heard English for between 25% and 75% of their language exposure and a second language for the remaining 75%–25%. These language exposure criteria are similar to other recent word learning studies with bilingual infants (Byers-Heinlein, Fennell, & Werker, 2013; Fennell et al., 2007). The average exposure to English was 48% and 51% for the 14- and 19-month-olds, respectively. Crucially, none of the infants' other languages used pitch contour contrastively. They were all nontonal languages, according to the *World Atlas of Language Structures* (Dryer & Haspelmath, 2011). In the 14-month-old group the languages were: Arabic (1), Bengali (1), French (2), Spanish (11), and Tongan (1). In the 19-month-old group the languages were: Amharic (1), French (1), Italian (3), Portuguese (1), and Spanish (10). Twelve additional infants were excluded from analyses because of fussiness or crying (7), moving out of the video frame (3), parental interference (1), or being distracted by an object during testing (1). One 19-month-old was identified as an outlier (looking time difference over 2.5 *SD* from the mean) and was excluded from analyses.

For Experiments 1 and 2, the demographics of the sample were: 38% White/Caucasian, 19% Mixed Race, 11% Black/African American, 2% Asian, 17% Other, and 13% not reported; 57% of the participants were Hispanic. All infants were tested between 2012 and 2014 in a small city in Northern California (Davis, CA). Infants were recruited from a database of families in the region who had expressed interest in participating in research.

Stimuli

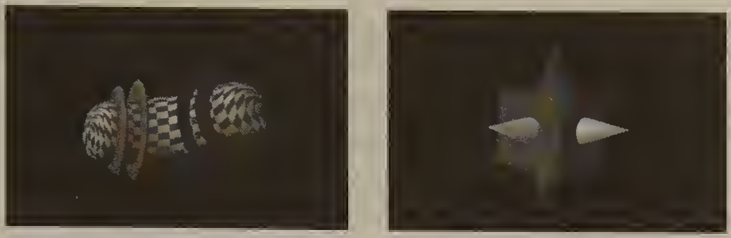
The stimuli were identical to the stimuli that Hay et al. (2015) designed for testing monolingual infants. The labels consisted of the syllable *ku* (/kʊ/) produced by a female native Mandarin Chinese speaker with a rising pitch contour (Mandarin Tone 2) and a falling pitch contour (Mandarin Tone 4). Figure 1 shows a spectrogram of each label with the fundamental frequency (F0, a measure of pitch) displayed. The voiced portion of the rising *ku* began with a frequency of 245 Hz, fell to 200 Hz over the first 130 ms of voicing, then rose to 290 Hz over the remainder of the syllable. The voiced portion of the falling *ku* began with a frequency of 320 Hz, which fell to 190 Hz over the remainder of the syllable. Total durations of the labels were similar (rising *ku* = 856 ms; falling *ku* = 867 ms). Tokens were repeated with 750 ms of silence separating them. Auditory stimuli were presented at 65 decibels (dB) of sound pressure level, the same intensity used by Hay et al. (2015).

As shown in Figure 2, the novel objects were three-dimensional images designed to differ in



Falling /kʊ/
Rising /kʊ/

Figure 1. Spectrogram and pitch contours of falling /kʊ/ (left) and rising /kʊ/ (right).



Object 1 Object 2

Figure 2. Objects used in label-object association task.

color, shape, and texture. Each object moved slowly from left to right across a black rectangle at the center of a television screen. The movement was not synchronized with the labels. Infants were randomly assigned to view Object 1 paired with rising *ku* and Object 2 paired with falling *ku*, or vice versa. The objects, along with their movement and presentation relative to the auditory signal, were identical to those used by Hay et al. (2015).

Procedure

The procedures, including the basic laboratory setup, the number of habituation and test trials, the trial duration, and the habituation criterion used here, were also the same as those used by Hay et al. (2015). Each infant was seated on a parent's lap in a sound-attenuated booth, approximately 1 m from a television with integrated speakers. A camera mounted below the screen was connected to an external monitor that allowed the experimenter to view the infant's responses. The program Habit X (Cohen, Atkinson, & Chaput, 2004) was used to present the audiovisual stimuli on the television and measure looking time. The experimenter indicated when the infant looked at the stimuli using a button press on the computer running Habit. To avoid bias, the parent listened to masking music on headphones and the experimenter was blind to the stimuli being presented.

The infant first viewed a pretest trial intended to familiarize him or her with the nature of the audiovisual task. During this trial, a novel object was paired with the syllable /la/ produced with a level pitch.

We used a modified version of the Switch task (Werker et al., 1998), which measures infants' ability to acquire novel associations between word forms and referents (see also Fennell et al., 2007; Graf Estes & Bowen, 2013; Werker et al., 2002). Each trial began with an animated cartoon clip to guide attention to the screen. Label learning took place during habituation trials. During each habituation trial, when the infant looked, a label–object pairing was presented. Each trial continued until the infant looked away for at least 1 s, or for a maximum of 20 s. The presentation of the two label–object pairings was randomized by block until the infant met the habituation criterion: Looking time across three consecutive trials (in a sliding window) decreased to 50% of the mean looking time across the first three trials. Using a sliding window and presenting the label–object pairs with the order randomized by block ensured that infants advanced to the test phase once they habituated, while maintaining very similar numbers of habituation trials (maximum difference of one) for each label–object pair.

The test trials immediately followed habituation. There were two types of test trials: same trials in which the original label-object associations were maintained and switch trials in which the familiar labels and objects occurred in new pairings. For example, during habituation and same trials Object 1 occurred with rising *ku*, and on switch trials Object 1 occurred with falling *ku*. There were two blocks of four test trials; each block contained two same trials and two switch trials. Infants were randomly assigned to participate in one of eight pseudo-randomized test orders that counterbalanced the presentation of same and switch trials. Preliminary tests indicated no effects of test order or test block therefore analyses collapsed across these variables. The rationale behind the task is that if infants learned the label–object pairings during habituation, they should look longer during switch test trials when those pairings are violated. Several previous experiments using the Switch task have presented only two test trials (one same trial, one switch trial; e.g., Fennell et al., 2007; MacKenzie, Graham, & Curtin, 2011; Werker et al., 1998). We used eight trials to maintain consistency with the method that Hay et al. (2015) used, which is also similar to several prior experiments (Graf Estes & Bowen, 2013; Graf Estes, Evans, Alibali, & Saffran, 2007; Hay, Pelucchi, Graf Estes, & Saffran, 2011).

Results and Discussion

The 14-month-olds met the habituation criterion following a mean of 12.6 trials (SD = 4.7), accumulating a total looking time during habituation of 137.2 s (SD = 61.2). The 19-month-olds met the habituation criterion following a mean of 9.3 trials (SD = 3.6, accumulating a total looking time during habituation of 106.4 s (SD = 56.8). All infants met the habituation criterion. Although the 19-month-olds habituated in fewer trials than the 14-month-olds, $t(30) = 2.25$, $p = .032$, $d = 0.80$, there was no significant difference in the total looking time during habituation, $t(30) = 1.47$, $p = .152$, $d = 0.52$.

To examine infants' learning of the labels, we performed a 2 (age: 14 months vs. 19 months; between subjects) × 2 (trial type: same vs. switch; within subjects) mixed analysis of variance (ANOVA) of infants' looking time. There was no main effect of age ($F < 1$) and no interaction of Age × Trial Type, ($F < 1$). There was a significant effect of trial type, $F(1,\ 30) = 10.77$, $p = .003$, $\eta_p^2 = .26$. Across age groups, infants looked longer during switch trials ($M = 8.6$ s, $SD = 3.3$) than same trials ($M = 7.0$ s, $SD = 2.4$).

The ANOVA did not reveal any difference in performance based on age. The pattern contrasts with Hay et al.'s (2015) recent findings using the same stimuli, in which 14-month-old monolinguals learned the labels, but 19-month-olds did not. To confirm that our results indeed reflected that infants at both ages learned the label–object associations, we performed paired samples *t* tests for each age group. As Figure 3 illustrates, infants in each age group looked significantly longer during switch trials compared to Same trials: 14 months, $t(15) = 2.28$, $p = .038$, $d = 0.63$, and 19 months, $t(15) = 2.57$, $p = .021$, $d = 0.52$. Thus, infants at 14 and 19 months of age reliably detected when the label–object pairings were switched, indicating that they successfully learned the new object labels.

The results described above suggest that bilingual infants displayed a different pattern of word learning than has been found in monolingual English-learning infants (Hay et al., 2015). To determine whether the bilingual pattern is reliably different from the monolingual pattern, we performed statistical comparisons with the data from Hay et al.'s (2015) work, which used the same procedures. At 14 months of age, a 2 (language background: monolingual vs. bilingual; between subjects) × 2 (trial type: same vs. switch; within subjects) mixed ANOVA of infants' looking time revealed no main effect of language background ($F < 1$) and no significant interaction ($F < 1$), but there was a significant effect of trial type, $F(1,\ 38) = 9.36$, $p = .004$, $\eta_p^2 = .20$. Across language backgrounds, 14-month-olds consistently displayed longer looking during switch trials. At 19 months of age, there was a different pattern. A 2 (language background) × 2 (trial type) mixed ANOVA of infants' looking time revealed no main effects of language Background ($F < 1$) or trial Type ($F < 1$). There was a significant Language background × Trial type interaction, $F(1,\ 38) = 4.41$, $p = .042$, $\eta_p^2 = .10$. The interaction confirms that 19-month-old monolinguals and bilinguals showed different looking patterns. Bilinguals tended to detect when the label–object pairings were switched, whereas monolinguals did not.

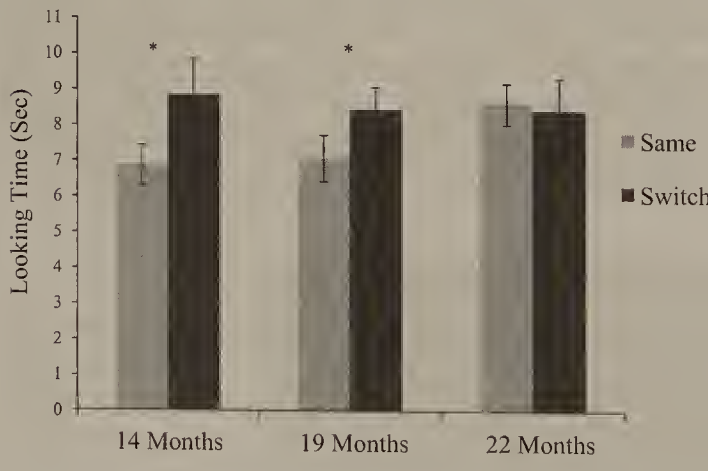


Figure 3. Mean looking time (in seconds) to same and switch test trials. Error bars represent standard errors. Asterisks indicate same versus switch trial looking times that are significantly different, $p < .05$.

Experiment 2

Experiment 1 indicated that bilingual infants did not show the same pattern of narrowing in their interpretation of pitch contour that monolinguals displayed in the same task and with the same stimuli (Hay et al., 2015). At 19 months of age, bilinguals seem to be more flexible than monolinguals in their interpretation of the sound variants that can differentiate words. In Experiment 2, we tested bilingual infants who have had several additional months of experience with their nontonal languages, 22-month-olds. We predicted that bilingual infants who are close to their second birthdays may have accumulated sufficient experience with how pitch is used in their native languages to rule it out as a means to differentiate words. Thus, by 22 months, they may show constraint in their interpretation of pitch contour in new words.

Method

Participants

The participants were 16 bilingual 22-month-old infants ($M = 22.5$ months, range = 22.0–23.5 months; 9 females). The infants met the same health and language inclusion criteria as in Experiment 1. The mean exposure to English was 55%. The nontonal second languages infants were exposed to were: French (3), Nepali (1) Spanish (11), and Tongan (1). Nine additional infants were excluded from analyses because of fussiness or crying (7) or parental interference (2).

Stimuli and Procedure

The stimuli and procedure were identical to Experiment 1.

Results and Discussion

Infants met the habituation criterion following a mean of 8.6 trials (*SD* = 3.1), accumulating a total looking time during habituation of 100.9 s (*SD* = 44.3). All infants met the habituation criterion. The 22-month-olds habituated in fewer trials than the 14-month-olds in Experiment 1, $t(30) = 2.82$, $p = .008$, $d = 1.02$. The 22-month-olds also had a marginally shorter total time to habituate, $t(30) = 1.19$, $p = .065$, $d = 0.69$. However, the 22-month-olds did not differ from the 19-month-olds in Experiment 1 in trials to habituate, $t(30) = 0.531$, $p = .599$, $d = 0.18$, or time to habituate, $t(30) = 0.307$, $p = .761$, $d = 0.11$. Interestingly, for all age groups, there were no significant correlations between trials or time to habituate and the magnitude of infants' preference for switch trials (all *p*s > .14). Thus, any differences in label learning performance cannot be attributed to differences in habituation.

To examine infants' learning, we performed a paired samples *t* test comparing looking time to same versus switch test trials. As shown in Figure 3, there was no significant difference in looking time, $t(15) = 0.229$, $p = .822$, $d = 0.07$. This suggests that the 22-month-old bilinguals did not learn the label–object pairings.

To examine whether the performance of the 22-month-olds was significantly different from the younger infants' performance in Experiment 1 (14- and 19-month-olds combined), we performed a 2 (experiment: Experiment 1 vs. Experiment 2; between subjects) × 2 (trial type: same vs. switch; within subjects) repeated measures ANOVA of infants' looking time. There were no main effects of experiment ($F < 1$) or trial type, $F(1, 46) = 2.52$, $p = .140$, $\eta_p^2 = .05$. However, there was a marginally significant Experiment × Trial Type interaction, $F(1, 46) = 3.77$, $p = .058$, $\eta_p^2 = .08$, suggesting that the bilingual 22-month-olds showed a weaker (*ns*) switch trial preference and weaker evidence of label learning than their younger bilingual peers.

General Discussion

We found that 14- and 19-month-old bilingual infants learned object labels that differed in a nonnative pitch contour contrast. They displayed flexibility in their interpretation of the types of sounds that act as words—young bilinguals treated distinct pitch contours as lexically contrastive even though they were not used contrastively in their own native languages. By 22 months of age, bilingual infants no longer displayed this flexibility. This pattern contrasts with the developmental trajectory of interpretive narrowing reported in previous research using the same methodology, in which monolingual English-learning 19-month-olds failed to learn labels with the same exact objects and pitch contours (Hay et al., 2015). Monolingual English-learning 14-month-olds succeeded in learning the labels, suggesting that accumulating experience with English led infants to restrict their interpretation of the sounds that can differentiate words in the middle of the 2nd year of life. Here, we present evidence that bilinguals show an extended period of flexibility in word learning. We suggest two hypotheses regarding why bilinguals maintain this interpretive flexibility.

First, bilinguals may maintain flexibility regarding possible word forms longer than monolinguals because they experience a protracted period of phonological development. They have two separate phonological systems to acquire, and hear less input in a given language than monolinguals do (Byers-Heinlein & Fennell, 2014; Costa & Sebastián-Gallés, 2014). The data are mixed regarding how bilingual experience affects the development of speech perception. Some experiments suggest that bilinguals have delays relative to monolinguals (Bosch & Sebastián-Gallés, 2003b; Ramon-Casas et al., 2009; Sebastián-Gallés & Bosch, 2009), whereas others do not (e.g., Burns et al., 2007; Sundara et al., 2008). The inconsistencies in the findings may be related to the languages and contrasts tested, as well as testing methodologies. When experimental tasks are designed to suit bilinguals' linguistic experiences, bilingual infants do not show delays relative to monolinguals (Albareda-Castellot et al., 2011; Mattock et al., 2010; see Sebastián-Gallés, 2011, for further discussion). Still, differences in performance across groups suggest that experience-driven language-specific speech perception may take longer to emerge for bilinguals than for monolinguals (Garcia-Sierra et al., 2011). Similarly, in word learning, bilinguals may entertain a broad range of options for how sounds make meaningful distinctions between words for a longer period than monolinguals do. Bilingual input contains greater variability in how words sound compared to monolinguals. Their two languages may divide up the acoustic-phonetic space differently, creating different phoneme boundaries and categories, as well as different phoneme combinations. Therefore, bilinguals have experience with a wide variety of ways that sounds can form words. Relative to monolinguals, phonological development in bilinguals may take longer to determine which sound variants differentiate words and which do not. Thus, in the present task, bilingual infants showed interpretive narrowing for pitch contour several months later than monolinguals.

A second hypothesis is broader in scope: Bilinguals may be more flexible than monolinguals in how they interpret linguistic input. In learning and processing two languages, bilinguals gain experience shifting their attention across the appropriate acoustic characteristics for detecting, recognizing, and comprehending words in each language. For example, infants learning Chinese and English must shift between a language that uses pitch contour to differentiate words and one that does not. French-English bilinguals must shift between English, which makes a phonemic distinction between θ and ð (as in "think" and "these," respectively) and French, which uses neither consonant. There is prior evidence that bilingual speech processing is exquisitely sensitive to the contexts in which phoneme contrasts are tested, even in infancy (Fennell & Byers-Heinlein, 2014; Mattock et al., 2010). Bilinguals may develop skills for shifting attention to the information that is relevant for processing a particular language when the input provides the necessary cues. These skills may apply broadly to novel linguistic input, not just to the infants' native languages. In the present experiment, the input indicated that pitch contour distinctions were important for differentiating words linked to objects. For 19-month-old monolingual infants, this experience was not sufficient to promote learning of the labels (Hay et al., 2015). At 19 months, bilingual infants may be able to use information present in the input to direct attention to what is meaningfully relevant in the moment. However, by 22 months, bilingual infants learning two nontonal languages develop constraints on tone processing. Their knowledge of how pitch is used in their languages is strong enough to override the cues that pitch is relevant in the current task. This does not mean that word learning is entirely unconstrained in younger bilinguals. For example, Fennell and Byers-Heinlein (2014) found that bilinguals were not more open than monolinguals to learning labels produced in accents that did not match their own language experience (i.e., bilinguals learned labels from bilingual-produced input, not monolingual-produced input). There may be further limits to their interpretive breadth. For example, we are testing whether the observed openness in bilingual word learning is limited to linguistic stimuli, or whether it applies to nonlinguistic sounds as well.

The idea that bilinguals have advantages over monolinguals in general perceptual and attentional processes has been supported in prior work. For example, bilingual infants can attend to visual cues in faces silently producing speech to differentiate between two languages, even when both languages are unfamiliar (Sebastián-Gallés, Albareda-Castellot, Weikum, & Werker, 2012). In contrast, monolinguals cannot. The perceptual and attentional characteristics of bilingual infants may reflect adaptations to their input (Costa & Sebastián-Gallés, 2014; Kovacs & Mehler, 2009). Perhaps protracted flexibility in word learning is another form of adaptation. Our proposal here is similar to Petitto et al.'s (2012) notion that bilingual experience produces increased demands on language processing, which effectively boosts the ability to analyze multiple dimensions of linguistic structure, "leaving open and agile linguistic processing in general" (p. 140).

One question that the present results raise is: Why do bilingual infants successfully learn labels that contrast in tones at ages that they have difficulty learning labels that contrast in native phonemes? Fennell et al. (2007) found that bilinguals failed to learn minimal pair–object labels at 17 months (but see Mattock et al., 2010, for alternative results). Monolinguals in the same task failed to learn the labels at 14 months, but succeeded at 17 months (Werker et al., 2002). For young monolinguals and for bilinguals, the difficulty may arise from determining which perceptible sound variants are relevant in their native languages. They may initially be over-inclusive regarding the interpretation of pitch contour contrasts, yet underattentive to small between-category phonetic differences. One possible contributing factor is the acoustic salience of the pitch contours that formed the tone contrast. The pitch contours unfolded over hundreds of milliseconds and moved across a wide frequency range. In comparison, the phonemes examined in similar experiments were typically stop consonants (Pater, Stager, & Werker, 2004; Werker et al., 2002), and thus differed in subtler characteristic than the tones, differing by only tens of milliseconds of voice onset time.

Furthermore, although neither of the bilingual infants' native languages used tone to contrast word meanings, pitch contour patterns are prevalent, salient, and significant in the infants' input. They are important for infants' interpretation of utterance boundaries and utterance types, as well as social information (Fernald, 1992; Seidl, 2007). Infants have substantial experience processing pitch and a great deal of input indicating that pitch is a crucial characteristic of the speech signal. The salience and significance of pitch are also likely to explain why 14-month-old monolinguals attend to tone contrasts during word learning in the Switch paradigm (Hay et al., 2015), but not nonlinguistic or noncommunicative sounds (MacKenzie et al., 2011). The present evidence indicates that bilinguals maintain this flexibility in their interpretation of pitch information for a longer period than monolinguals.

An additional consideration is that the different developmental patterns of monolingual and bilingual infants may not actually be due to bilingualism, as we have proposed. Rather, early interpretive narrowing could occur specifically in English learners and the infants in the present experiment could show relative flexibility in label learning because they are learning a language other than English. This explanation seems unlikely for two reasons. First, none of the infants in this experiment were exposed to tonal languages, so they had no prior experience with the contrastive use of pitch contours. For example, consider Spanish and French, which were the most common second languages in our sample. In Spanish, like in English, pitch is a component of lexical stress, and lexical stress patterns can differentiate between words (Beckman, Díaz-Campos, Tevis McGory, & Morgan, 2002). In French, stress is found at the phrase level, rather than the lexical level, and is realized through final lengthening with no increase in pitch or intensity (Vaissiere, 1991). However, in all three of these languages, pitch is part of the signal conveying information such as emphasis, emotional tone, speaking register, and sentence context (e.g., question vs. declarative statement). All of this information affects the interpretation of utterances, but does not provide information indicating that pitch contour differences alone should differentiate between words. Second, the infants were exposed to a wide range of languages. Generally, languages manipulate pitch in different ways, as evidenced by the differences between Spanish and French. Thus, it seems improbable that such a broad group of languages would have the same effect on label learning, namely, to support the acceptance of pitch contour as lexically contrastive. Rather, the observed developmental pattern seems to be due to the experience of becoming bilingual. Future research will be necessary to determine how general the observed effect is, given that there are many different forms that bilingualism can take. For example, it is not yet clear whether our effect is specific to infants who are simultaneously exposed to two languages, or whether it also occurs for sequential bilinguals.

Singh et al. (2014) also found evidence of interpretive narrowing in bilingual infants' treatment of pitch contour; they found that English monolinguals and nontonal language bilinguals showed the same developmental trajectory, whereas we found different developmental patterns across groups. Specifically, Singh et al. found that both monolingual and bilingual 18-month-olds detected changes in the pitch contours of new words. By 24 months, infants in both groups showed constraint in their interpretation of pitch and did not attend to pitch contour changes. In the present experiments, we found that bilinguals showed successful learning of labels that differed only in pitch contour at 14 and 19 months, but not at 22 months, whereas Hay et al. (2015) found that English monolinguals in the same task learned the labels at 14 months, but not at 19 months.

As discussed in the Introduction, we propose that the difference in findings for monolinguals' processing of pitch contour in the middle of the 2nd year arose primarily because of differences in the designs of the tasks used by Hay et al. (2015) versus Singh et al. (2014). Differences in task designs may also explain the patterns of bilinguals' performance across studies. Singh et al.'s (2014) task incorporated substantial referential support for word learning. The labels were presented in carrier phrases and infants viewed supplementary trials with familiar objects. Both of these characteristics have been shown to promote attention to detail in the sounds of words during label learning (Fennell & Waxman, 2010). In contrast, studies using the Switch paradigm (similar to the present experiments), without referential support, have demonstrated that 14-month-olds have difficulty attending to phonetic detail in new words (Pater et al., 2004; Stager & Werker, 1997). The apparently simple Switch task uncovered a vulnerability in processing phonetic detail that was not apparent in richer tasks.

The importance of task demands in revealing developmental patterns is supported in the PRIMIR model of infant speech perception and word learning (Curtin et al., 2011; Werker & Curtin, 2005; Yoshida, Fennell, Swingley, & Werker, 2009). For monolinguals, the differences in methodology across Hay et al.'s (2015) and Singh et al.'s (2014) studies revealed an unstable characteristic of infants' processing. In the middle of the 2nd year, the methods by which object labels are introduced and tested affects how infants direct attention to the sounds of the words. Referential information may free cognitive resources, allowing infants to attend more closely to the acoustic-phonetic details of new words. In Singh et al.'s task, the referential context may have promoted monolingual and bilingual 18-month-olds' attention to detail, even for a dimension of speech that is not used to differentiate word meanings in their native languages. Seventeen- to 19-month-old monolingual English-learning infants do not show this level of attention to pitch contour in tasks that lack referential support (Hay et al., 2015). Similarly, May and Werker (2014) found that monolingual 14-month-olds, as well as 20-month-olds with small vocabularies, learned labels containing non-native click contrasts in a referential context, but not without referential cues. Twenty-month-olds with larger vocabularies did not learn the labels in either context.

The present findings indicate that bilingual 19-month-olds do not require referential support to attend to pitch contour when learning new words. We propose that this indicates that for bilinguals around 18–19 months of age, attention to pitch contour in new words is robust; it occurs in both word learning tasks in which referential support is present and in tasks in which it is absent. By 22 months of age, bilinguals no longer show attention to pitch in the Switch paradigm when learning labels that differ only in pitch, and at 24 months they no longer show attention to pitch in the referential paradigm (Singh et al., 2014). Thus, by around age 2, knowledge of how pitch is used in bilinguals' native languages may lead to the development of firm constraints on how pitch is interpreted, constraints that are not readily perturbed by variation in learning and testing environments.

We expect that there is a convergence of factors that affect infants' acceptance of novel word forms, including the availability of referential cues, the details of the sound sequences, as well as the infant's experience and developmental level. For labels consisting of native sounds, the presence of referential information may promote earlier sensitivity to phoneme contrasts than without referential cues (Fennell & Waxman, 2010; Werker et al., 2002). For labels containing non-native sounds, referential information may also support greater attention to detail, including attention to sound contrasts that are not lexically contrastive in the native language(s). In the absence of referential cues, learning appears more constrained (Hay et al., 2015; May & Werker, 2014; Singh et al., 2014). In the present work, the absence of referential cues also revealed extended flexibility in label learning for bilingual infants. A key point that merits further attention is the generality of bilinguals' extended flexibility in label learning. We have proposed that infants maintain attention to pitch contour in labels because of its salience and prevalence in the input. It will be important to expand the investigation to other non-native sound contrasts varying in salience and degree of similarity to the native input. Some non-native contrasts may be so distant from native-language experience that infants do not readily entertain them as labels. Bilinguals may show a developmental progression similar to monolinguals in these cases, or they may be open to some types of sound contrasts that monolinguals are not open to. These ideas remain to be tested.

In conclusion, the current results indicate that bilingual infants are flexible in their interpretation of the sounds that make meaningful distinctions between words for a longer developmental period than monolinguals infants are. This flexibility seems to represent an appropriate strategy for bilingual infants. It fits with their phonological and lexical experience and the demands of their linguistic environments.

References

Albareda-Castellot, B., Pons, F., & Sebastián-Gallés, N. (2011). The acquisition of phonetic categories in bilingual infants: New data from an anticipatory eye movement paradigm. *Developmental Science*, *14*, 395–401. doi:10.1111/j.1467-7687.2010.00989.x

Beckman, M. E., Díaz-Campos, M., Tevis McGory, J., & Morgan, T. A. (2002). Intonation across Spanish, in the Tones and Break Indices framework. *Probus*, *14*, 9–36. doi:10.1515/prbs.2002.008

Bialystok, E., Barac, R., Blaye, A., & Poulin-Dubois, D. (2010). Word mapping and executive functioning in young monolingual and bilingual children. *Journal of Cognition and Development*, *11*, 485–508. doi:10.1080/15248372.2010.516420

Bialystok, E., Craik, F., & Luk, G. (2008). Cognitive control and lexical access in younger and older bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 859–873. doi:10.1037/0278-7393.34.4.859

Bosch, L., & Ramon-Casas, M. (2011). Variability in vowel production by bilingual speakers: Can input properties hinder the early stabilization of contrastive categories? *Journal of Phonetics, 39,* 514–526. doi:10.1016/j.wocn.2011.02.001

Bosch, L., & Sebastián-Gallés, N. (2003a). *Language experience and the perception of a voicing contrast in fricatives: Infant and adult data.* Paper presented at the 15th International Congress of the Phonetic Sciences, Barcelona, Spain.

Bosch, L., & Sebastián-Gallés, N. (2003b). Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life. *Language and Speech, 46,* 217–243. doi:10.1177/00238309030460020801

Burns, T. C., Yoshida, K. A., Hill, K., & Werker, J. F. (2007). The development of phonetic representation in bilingual and monolingual infants. *Applied Psycholinguistics, 28,* 455–474. doi:10.1017/s0142716407070257

Byers-Heinlein, K. (2013). Parental language mixing: Its measurement and the relation of mixed input to young bilingual children's vocabulary size. *Bilingualism: Language and Cognition, 16,* 32–48. doi:10.1017/S1366728912000120

Byers-Heinlein, K., & Fennell, C. T. (2014). Perceptual narrowing in the context of increased variation: Insights from bilingual infants. *Developmental Psychobiology, 56,* 274–291. doi:10.1002/dev.21167

Byers-Heinlein, K., Fennell, C. T., & Werker, J. F. (2013). The development of associative word learning in monolingual and bilingual infants. *Bilingualism: Language and Cognition, 16,* 198–205. doi:10.1017/S1366728912000417

Byers-Heinlein, K., & Werker, J. F. (2013). Lexicon structure and the disambiguation of novel words: Evidence from bilingual infants. *Cognition, 128,* 407–416. doi:10.1016/j.cognition.2013.05.010

Cohen, L. B., Atkinson, D. J., & Chaput, H. J. (2004). *Habit X: A new program for obtaining and organizing data in infant perception and cognition studies (Version 1.0).* Austin: University of Texas.

Costa, A., & Sebastián-Gallés, N. (2014). How does the bilingual experience sculpt the brain? *Nature Reviews Neuroscience, 15,* 336–345. doi:10.1038/nrn3709

Curtin, S., Byers-Heinlein, K., & Werker, J. F. (2011). Bilingual beginnings as a lens for theory development: PRIMIR in focus. *Journal of Phonetics, 39,* 492–504. doi:10.1016/j.wocn.2010.12.002

Dryer, M. S., & Haspelmath, M. (2011). *Word atlas of language structures.* Munich, Germany: Max Planck Digital Library. Retrieved from http://wals.info/

Fennell, C. T., & Byers-Heinlein, K. (2014). You sound like mommy: Bilingual and monolingual infants learn words best from speakers typical of their language environments. *International Journal of Behavioral Development, 38,* 309–316. doi:10.1177/0165025414530631

Fennell, C. T., Byers-Heinlein, K., & Werker, J. F. (2007). Using speech sounds to guide word learning: The case of bilingual infants. *Child Development, 78,* 1510–1525. doi:10.1111/j.1467-8624.2007.01080.x

Fennell, C. T., & Waxman, S. R. (2010). What paradox? Referential cues allow for infant use of phonetic detail in word learning. *Child Development, 81,* 1376–1383. doi:10.1111/j.1467-8624.2010.01479.x

Fernald, A. (1992). Human maternal vocalizations to infants as biologically relevant signals: An evolutionary perspective. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 391–428). New York, NY: Oxford University Press.

Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech, 1,* 126–152. doi:10.1177/002383095800100207

Garcia-Sierra, A., Rivera-Gaxiola, M., Percaccio, C. R., Conboy, B. T., Romo, H., Klarman, L., . . . Kuhl, P. K. (2011). Bilingual language learning: An ERP study relating early brain responses to speech, language input, and later word production. *Journal of Phonetics, 39,* 546–557. doi:10.1016/j.wocn.2011.07.002

Graf Estes, K., & Bowen, S. (2013). Learning about sounds contributes to learning about words: Effects of prosody and phonotactics on infant word learning. *Journal of Experimental Child Psychology, 114,* 405–417. doi:10.1016/j.jecp.2012.10.002

Graf Estes, K., Edwards, J., & Saffran, J. R. (2011). Phonotactic constraints on infant word learning. *Infancy, 16,* 180–197. doi:10.1111/j.1532-7078.2010.00046.x

Graf Estes, K., Evans, J. L., Alibali, M. W., & Saffran, J. R. (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychological Science, 18,* 254-260. doi:10.1111/j.1467-9280.2007.01885.x

Grosjean, F. (2010). *Bilingualism: Life and reality.* Cambridge, MA: Harvard University Press.

Gussenhoven, C. (2004). *The phonology of tone and intonation.* Cambridge, UK: Cambridge University Press.

Hay, J. F., Graf Estes, K., Wang, T., & Saffran, J. R. (2015). From flexibility to constraint: The contrastive use of lexical tone in early word learning. *Child Development, 86,* 10–22. doi:10.1111/cdev.12269

Hay, J. F., Pelucchi, B., Graf Estes, K., & Saffran, J. R. (2011). Linking sounds to meanings: Infant statistical learning in a natural language. *Cognitive Psychology, 63,* 93–106. doi:10.1016/j.cogpsych.2011.06.002

Hoff, E., Core, C., Place, S., Rumiche, R., Señor, M., & Parra, M. (2012). Dual language exposure and early bilingual development. *Journal of Child Language, 39,* 1–27. doi:10.1017/S0305000910000759

Kovacs, Å. M., & Mehler, J. (2009). Cognitive gains in 7-month-old bilingual infants. *Proceedings of the National Academy of Sciences of the United States of America, 106,* 6556–6560. doi:10.1073/pnas.0811323106

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and

12 months. *Developmental Science, 9,* F13–F21. doi:10.1111/j.1467-7687.2006.00468.x

MacKenzie, H., Graham, S. A., & Curtin, S. (2011). 12-month-olds priviledge words over other linguistic sounds in an associative learning task. *Developmental Science, 14,* 249–255. doi:10.1111/j.1467-8624.2012.01764.x

Mattock, K., & Burnham, D. (2006). Chinese and English infants' tone perception: Evidence for perceptual reorganization. *Infancy, 10,* 241–265. doi:10.1207/s15327078in1003_3

Mattock, K., Molnar, M., Polka, L., & Burnham, D. (2008). The developmental course of lexical tone perception in the first year of life. *Cognition, 106,* 1367–1381. doi:10.1016/j.cognition.2007.07.002

Mattock, K., Polka, L., Rvachew, S., & Krehm, M. (2010). The first steps in word learning are easier when the shoes fit: Comparing monolingual and bilingual infants. *Developmental Science, 13,* 229–243. doi:10.1111/j.1467-7687.2009.00891.x

May, L., & Werker, J. F. (2014). Can a click be a word? Infants' learning of non-native words. *Infancy, 19,* 281–300. doi:10.1111/infa.12048

Murray, I. R., & Arnott, J. L. (1993). Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *Journal of the Acoustical Society of America, 93,* 1097–1108. doi:10.1121/1.405558

Namy, L. L. (2001). What's in a name when it isn't a word? 17-month-olds' mapping of nonverbal symbols to object categories. *Infancy, 2,* 73–86. doi:10.1207/S15327078IN0201_5

Oller, D. K., Eilers, R. E., Urbano, R., & Cobo-Lewis, A. B. (1997). Development of precursors to speech in infants exposed to two languages. *Journal of Child Language, 24,* 407–425. doi:10.1017/S0305000997003097

Pater, J., Stager, C. L., & Werker, J. F. (2004). The lexical acquisition of phonological contrasts. *Language, 80,* 361–379. doi:10.1353/lan.2004.0141

Pearson, B. Z., Fernandez, S. C., & Oller, D. K. (1993). Lexical development in bilingual infants and toddlers: Comparison to monolingual norms. *Language Learning, 43,* 93–120. doi:10.1111/j.1467-1770.1993.tb00174.x

Petitto, L. A., Berens, M. S., Kovelman, I., Dubins, M. H., Jasinska, K., & Shalinsky, M. (2012). The "perceptual wedge hypothesis" as the basis for bilingual babies' phonetic processing advantage: New insights from fNIRS brain imaging. *Brain and Language, 121,* 130–143. doi:10.1016/j.bandl.2011.05.003

Petitto, L. A., Katerelos, M., Levy, B. G., Gauna, K., Tétreault, K., & Ferraro, V. (2001). Bilingual signed and spoken language acquisition from birth: Implications for the mechanisms underlying early bilingual language acquisition. *Journal of Child Language, 28,* 453–496. doi:10.1017/S0305000901004718

Poulin-Dubois, D., Blaye, A., Coutya, J., & Bialystok, E. (2011). The effects of bilingualism on toddlers' executive functioning. *Journal of Experimental Child Psychology, 108,* 567–579. doi:10.1016/j.jecp.2010.10.009

Ramon-Casas, M., Swingley, D., Sebastián-Gallés, N., & Bosch, L. (2009). Vowel categorization during word recognition in bilingual toddlers. *Cognitive Psychology, 59,* 96–121. doi:10.1016/j.cogpsych.2009.02.002

Rivera-Gaxiola, M., Silva-Pereyra, J., & Kuhl, P. K. (2005). Brain potentials to native and non-native speech contrasts in 7- and 11-month-old American infants. *Developmental Science, 8,* 162–172. doi:10.1111/j.1467-7687.2005.00403.x

Scott, L. S., Pascalis, O., & Nelson, C. A. (2007). A domain-general theory of the development of perceptual discrimination. *Current Directions in Psychological Science, 16,* 197–201. doi:10.1111/j.1467-8721.2007.00503.x

Sebastián-Gallés, N. (2011). Bilingual language acquisition: Where does the difference lie? *Human Development, 53,* 245–255. doi:10.1159/000321282

Sebastián-Gallés, N., Albareda-Castellot, B., Weikum, W. M., & Werker, J. F. (2012). A bilingual advantage in visual language discrimination in infancy. *Psychological Science, 23,* 994–999. doi:10.1177/0956797612436817

Sebastián-Gallés, N., & Bosch, L. (2009). Developmental shift in the discrimination of vowel contrasts in bilingual infants: Is the distributional account all there is to it? *Developmental Science, 12,* 874–887. doi:10.1111/j.1467-7687.2009.00829.x

Seidl, A. (2007). Infants' use and weighting of prosodic cues in clause segmentation. *Journal of Memory and Language, 57,* 24–48. doi:10.1016/j.jml.2006.10.004

Singh, L., Hui, T. J., Chan, C., & Golinkoff, R. M. (2014). Influences of vowel and tone variation on emergent word knowledge: A cross-linguistic investigation. *Developmental Science, 17,* 94–109. doi:10.1111/desc.12097

So, C. K., & Best, C. T. (2010). Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. *Language and Speech, 53,* 273–293. doi:10.1177/0023830909357156

Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature, 388,* 381–382. doi:10.1038/41102

Sundara, M., Polka, L., & Baum, S. (2006). Production of coronal stops by simultaneous bilingual adults. *Bilingualism: Language and Cognition, 9,* 97–114. doi:10.1017/S1366728905002403

Sundara, M., Polka, L., & Molnar, M. (2008). Development of coronal stop perception: Bilingual infants keep pace with their monolingual peers. *Cognition, 108,* 232–242. doi:10.1016/j.cognition.2007.12.013

Vaissiere, J. (1991). Rhythm, accentuation and final lengthening in French. In J. Sundberg, L. Nord, & R. Carlson (Eds.), *Music, language, speech and brain* (Vol. 59, pp. 108–120). Basinstoke, UK: Macmillan.

Werker, J. F., Cohen, L. B., Lloyd, V. L., Casasola, M., & Stager, C. L. (1998). Acquisition of word-object associations by 14-month-old infants. *Developmental*

Psychology, 34, 1289–1309. doi:10.1037/0012-1649.34.6.1289

Werker, J. F., & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development, 1,* 197–234. doi:10.1207/s15473341lld0102_4

Werker, J. F., Fennell, C. T., Corcoran, K. M., & Stager, C. L. (2002). Infants' ability to learn phonetically similar words: Effects of age and vocabulary size. *Infancy, 3,* 1–30. doi:10.1207/15250000252828226

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7,* 49–63. doi:10.1016/s0163-6383(84)80022-3

Woodward, A. L., & Hoyne, K. L. (1999). Infants' learning about words and sounds in relation to objects. *Child Development, 70,* 65–77. doi:10.1111/1467-8624.00006

Yeung, H. H., Chen, K. H., & Werker, J. F. (2013). When does native language input affect phonetic perception? The precocious case of lexical tone. *Journal of Memory and Language, 68,* 123–139. doi:10.1016/j.jml.2012.09.004

Yip, M. (2002). *Tone.* Cambridge, UK: Cambridge University Press.

Yoshida, K. A., Fennell, C. T., Swingley, D., & Werker, J. F. (2009). Fourteen-month-old infants learn similar-sounding words. *Developmental Science, 12,* 412–418. doi:10.1111/j.1467-7687.2008.00789.x

Child Development, September/October 2015, Volume 86, Number 5, Pages 1386–1405

Prelinguistic Relational Concepts: Investigating Analogical Processing in Infants

Alissa L. Ferry, Susan J. Hespos, and Dedre Gentner
Northwestern University

This research asks whether analogical processing ability is present in human infants, using the simplest and most basic relation—the *same–different* relation. Experiment 1 ($N = 26$) tested whether 7- and 9-month-olds spontaneously detect and generalize these relations from a single example, as previous research has suggested. The attempted replication failed. Experiment 2 asked whether infants could abstract the relation via analogical processing (Experiment 2, $N = 64$). Indeed, with four exemplars, 7- and 9-month-olds could abstract the *same–different* relation and generalize it to novel pairs. Furthermore, prior experience with the objects disrupted learning. Facilitation from multiple exemplars and disruption by individual object salience are signatures of analogical learning. These results indicate that analogical ability is present by 7 months.

Analogical ability—the ability to make relational comparisons between objects, events, or ideas, and to think about relations independently of a particular set of arguments—is a cornerstone of higher reasoning abilities. Learning by analogy is a powerful and efficient method for acquiring new information. Equally important, analogical comparison facilitates the formation of abstract categories and rules (Gentner & Medina, 1998; Gick & Holyoak, 1983). Indeed, recent theoretical perspectives have asserted that analogical ability is the key capacity supporting higher order cognition and differentiating human cognitive capacity from that of other primates (Gentner, 2003, 2010; Penn, Holyoak, & Povinelli, 2008).

Our goal in this research is to trace the development of relational processing—specifically, to discover whether processes of analogical learning are present in prelinguistic human infants. We focus on the *same–different* relation, for four reasons. First, it is arguably the simplest and most basic relation, and is likely to be available early in development. Second, the perception of sameness is critical to making sense of the world; in the words of William James (1890), the "*sense of sameness* is the very keel and backbone of our thinking" (p. 459; quoted in Wasserman & Young, 2010). A third point is that the *same–different* relation has the advantage that it can be tested by a simple perceptual task: the two-item same–different task, which is readily adaptable for infant studies. If a participant can discriminate between *same* pairs {AA, BB, CC. . .} and *different* pairs {AB, CD, EF. . .} and can generalize this distinction to new exemplars (XX vs. YZ), then we can infer that they have represented abstract *same* and *different* relations.

The final reason for focusing on the *same–different* relation is that it has been the main focus of comparative research on relational ability in nonhuman species. Beginning with Premack's (1983) seminal paper, which argued that chimpanzees (but not monkeys) possess significant relational ability, much of this research has focused on the *same–different* relation (Fagot & Thompson, 2011; Flemming,

---

Alissa L. Ferry is currently at Scuola Internazionale Superiore di Studi Avanzati.

This research was supported by NSF Grant BCS-1423917 to Susan J. Hespos, NSF SLC Grant SBE-0541957 awarded to the Spatial Intelligence and Learning Center (SILC), and by ONR Grant N00014-92-J-1098 to Dedre Gentner. Preparation of the article was supported in part by the European Research Council under the European Unions' Seventh Framework Programme (FP7/2007-2013)/European Research Council Grant Agreement 269502 (PASCAL), which supported Alissa Ferry during manuscript preparation and an award to Dedre Gentner from the Alexander von Humboldt Foundation and by a fellowship at the Hanse-Wissenschaftskolleg, Delmenhorst, Germany. We are indebted to parents who agreed to have their infants participate in the research. We thank members of the Infant Cognition Lab for help with data collection. Special thanks go to Don Tyrrell for his willingness to re-engage with these topics after 18 years of retirement. His thoughtful discussion has enhanced the quality of this work and has confirmed that these ideas retain their relevance across several decades. Finally, we thank Susan Carey for many insightful comments on this work.

Correspondence concerning this article should be addressed to Alissa L. Ferry, Cognitive Neuroscience Sector, Scuola Internazionale Superiore di Studi Avanzati, Via Bonomea 265, 34136 Trieste, Italy. Electronic mail may be sent to aferry@sissa.it.

© 2015 The Authors
Child Development © 2015 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2015/8605-0006
DOI: 10.1111/cdev.12381

Beran, & Washburn, 2007; Giurfa, Zhang, Jennett, Menzel, & Srinivasan, 2001; Thompson, Oden, & Boysen, 1997; Wasserman & Young, 2010). This research has shown that although many species, including pigeons, macaques, and honeybees, can pass the object match-to-sample test (given a standard A, choose the matching alternative, A, over B), the ability to pass the *same–different* task (making one response for AA, BB, CC, etc. and another for AB, CD, EF, etc.) is sharply graded across species. For example, Wright and Katz (2006) found that rhesus monkeys and capuchin monkeys required over 4,700 trials to learn to distinguish *same* pairs from *different* pairs and generalize the distinction to novel pairs. These findings show the difficulty of *same–different* relational matching for monkeys. In contrast, chimpanzees can learn and generalize the *same–different* relation in 150–500 trials (Premack, 1983; Robinson, 1955) and human 3-year-olds given a sorting task can learn to distinguish *same* and *different* relations after fewer than 24 trials (Christie & Gentner, 2014). A more challenging task, in which the subject must match a *same* pair to either a *same* (correct) or a *different* pair, was given to baboons (Fagot & Thompson, 2011). Only 6 of 29 animals were able to reach criterion, and they required a mean of 21,000 trials. In contrast, human 4-year-olds succeed at this task (for the *same* relation) in just eight trials without feedback (Christie & Gentner, 2014). Thus, relational ability—even with the basic *same–different* relation—varies strongly across primate species.

The Ontogeny of Human Analogical Ability

Adult humans—and even children—are virtuosos of relational ability compared with other primates. But how does this ability develop?

How do infants come to be able to encode and generalize relations between things? We can distinguish three broad positions. The strongest possibility is that human infants are born with a set of privileged relations that they naturally use to encode experience—a set of relational semantic primitives. Such a set would almost certainly include the *same–different* relation. In this case, infants would be able to encode abstract *same* and *different* relations even from a single exemplar. A second possibility is that human infants are born with analogical ability, with which they can learn abstract relations from experience, but without a starting set of abstract relations. In this case, infants would be unable to spontaneously encode the *same* and *different* relations from a single exemplar but might be able to use analogical learning mechanisms to encode these relations by comparing exemplars. A third possibility is that analogical ability is not inherent in human biology, but develops by combining other abilities through cultural and linguistic experience. This possibility receives some support from studies that show that learning and using relational language facilitates relational insight in children (Christie & Gentner, 2014; Gentner, 2003, 2010; Kotovsky & Gentner, 1996). In this case, we would not expect to see evidence of abstract *same* and *different* relations until children have acquired a degree of cultural and linguistic experience—perhaps including some knowledge of relational terms.

The first and strongest possibility is that *same* and *different* are part of an innate "starter kit" of abstract relational primitives, akin to the semantic features postulated by some early theories of meaning acquisition (e.g., Clark, 1973). This account also assumes that infants are born with analogical ability—that is, the ability to align and map between relations. We consider this possibility in Experiment 1, and then turn to the other two possibilities.

The idea that abstract *same* and *different* relations are present from the start has had wide influence, particularly in comparative psychology, through the pioneering work of Tyrrell, Stauffer, and Snowman (1991). This research suggested that 7-month-old humans spontaneously encode and transfer *same* and *different* relations from a single exemplar—implying that human infants can encode abstract *same–different* relation without training. Because of the importance of these claims, our first order of business was to examine and replicate this study. Unfortunately (to preview) our attempted replication failed; further, inspection of the paper revealed that it does not support the strong possibility above.

We first review this study and then present our attempted replication. Tyrrell et al. (1991) tested twenty-two 7-month-old infants. Half of the infants were familiarized to an exemplar of the *same* relation (AA; presented twice); the other half were familiarized to an exemplar of the *different* relation (BC), also presented twice. In the first test trial, infants were shown a side-by-side presentation of AA and BC—that is, the pair they had seen along with a pair they had not seen (and which depicted a different relation from the familiar pair)—until 10 s of looking at either pair accumulated. In the second test trial, they saw two new pairs (DD and EF), and the relative positions of the *same* and *different* pairs were switched. The reasoning was if the infants showed a novelty preference for the relation that they had not been familiarized to, this would indicate that they already possess an abstract *same–different* relation, which can be generalized from a single exemplar. Combining both of these test trials, Tyrrell et al. (1991) found just this pattern: After habituation to *same* the infants looked significantly longer at the novel relation (*different*) in the test trials than at the familiar relation (*same*). The same novelty preference was also found for infants habituated to *different*. Because the infants had seen only one instance of the relation prior to testing, this seemed to suggest that 7-month-olds spontaneously encode abstract relational concepts—potentially a very important result. Given the importance of this question, and the oft-cited interpretation that their study represents evidence of generalizing *same–different* relations from a single exemplar, we set out to replicate the experiment using a single exemplar, to test whether infants can spontaneously detect the *same–different* relations.

Before turning to the replication, we note that the original Tyrrell et al. (1991) publication contains some ambiguities regarding the exact method. However, subsequent discussions of this influential work have adopted the interpretation offered here —namely, that infants generalized *same* (or *different*) from a single example—indicating that they already possessed abstract representations of *same* and *different* (e.g., Dewar & Xu, 2010; Marcus, Fernandes, & Johnson, 2007; Thompson & Oden, 1995, 2000; Tyrrell, Zingardo, & Minard, 1993; Zentall, Wasserman, Lazareva, Thompson, & Rattermann, 2008).

Experiment 1

In this study, we set out to replicate the research of Tyrrell et al. (1991). Our goal was to discover whether 7-month-old infants would encode an abstract *same* (or *different*) relation from a single familiarization exemplar. We also tested a group of 9-month-old infants to capture potential developmental differences, as previous research had suggested that by this age, infants show some evidence of understanding *same* and *different* in a different task (Addyman & Mareschal, 2010). We followed an interpretation of Tyrrell et al.'s experimental design with only a single familiarization exemplar. As shown in Figure 1A, during familiarization, infants saw two presentations of the *same* example pair (e.g., AA). In the first test trial (old), they chose between this initial example pair (AA) and a new pair depicting the other relation (e.g., BC). In the second test trial (novel), *same* and *different* were instantiated with new objects. Only if infants differentiate the relations on the second trial can we infer that they already possess abstract *same–different* relations.

Method

Participants

The participants were 26 healthy, full-term infants, ten 7-month-olds (5 male and 5 female) ranging from 6 months 23 days to 7 months

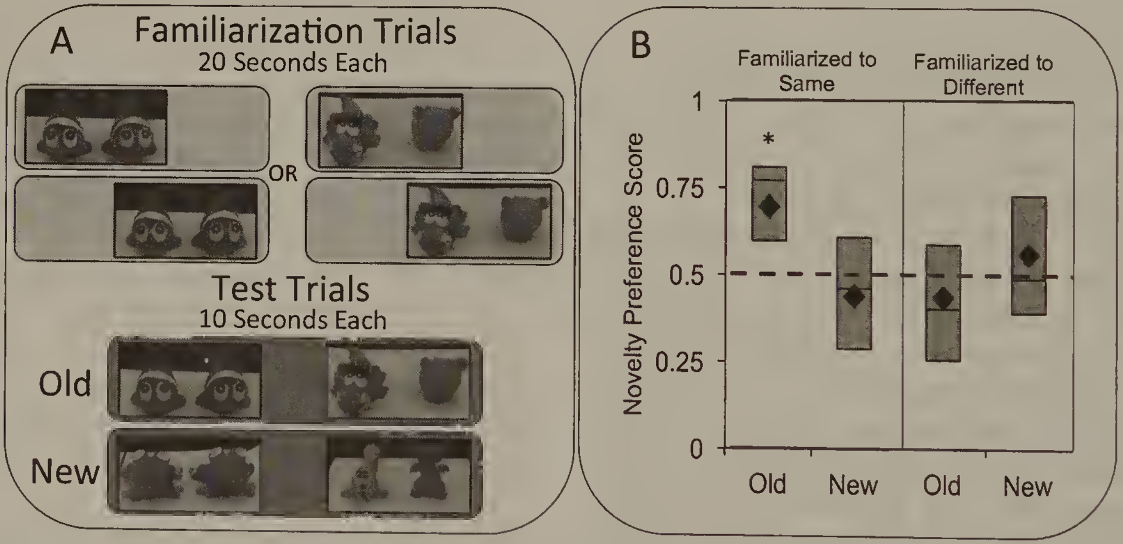


Figure 1. (A) Schematic of Experiment 1. During familiarization trials, infants were shown either a pair of *same* or a pair of *different* objects. During the test trials, infant saw two pairs of objects presented side by side. For the old test trial, one pair was identical to the pair seen during familiarization; the other pair consisted of novel objects and a novel relation. For the new test trial, all the objects were novel and one pair depicted the familiarized relation and the other depicted the novel relation. (B) Test trial novelty preference scores in Experiment 1. Box plots showing mean preference score for the novel relation over the familiar relation during both test trials. Black diamonds represent means, the central line in each box is the median, and the upper and lower portions of each box represent the third and first quartiles, respectively. The dashed line represents chance performance.
$^*p < .05$.

24 days (*M* = 7 months 4 days) and sixteen 9-month-olds (6 male and 10 female) ranging from 8 months 19 days to 9 months 16 days (*M* = 8 months 29 days). Approximately half the infants were assigned to the *same* condition (*n* = 11); the other half, to the *different* condition (*n* = 15). Four additional infants were tested but eliminated from the final analysis, three because of parental interference and one because he looked only at the right side of the stage during every test trial.

For this and the following experiment, we obtained infants' names from commercial mailing lists. Parents were contacted by letters and provided informed consent before the experiment. They were given $20 as compensation. The ethnicity of the sample was 81% non-Hispanic, 14% Hispanic, and the remaining 5% chose not to answer. The racial makeup was 76% white, 6% Black/African American, 4% Asian, 1% Native Hawaiian/Pacific Islander, 1% American Indian/Alaskan Native, 7% multiracial. The remaining 5% chose not to answer. The mothers' highest level of education included 1% some high school, 6% high school diploma, 14% some college, and 75% college degree or higher; the remaining 4% chose not to answer. The fathers' highest level of education included 1% some high school, 4% high school diploma, 11% some college, 80% college degree or higher; the remaining 4% chose not to answer.

Apparatus

Parents sat in a chair with infants on their lap facing a wooden puppet stage that displayed all stimulus objects. The stage measured 243.5 cm high, 128 cm wide, and 61 cm deep. The opening in the front of the stage that displayed the objects was 93 cm above the floor, 61 cm high, and 106 cm wide. The back wall had two rectangular openings with cloth fringe over the openings that allowed the experimenter to manipulate the objects between trials. A screen that covered the infants' view of the stage was raised and lowered between trials (see Figure 1A).

The stimuli consisted of four pairs of objects, each attached by Velcro to a 26.5 cm × 15.5 cm piece of cardboard covered with contact paper. The items in each pair were approximately 12.5 cm apart from each other. Two pairs of objects instantiated the *same* relation (two identical fish and two identical frogs) and two pairs of objects instantiated the *different* relation (a crocodile and a pig paired together and a dog and a dragon paired together).

Events

Familiarization trials. When the screen went up at the start of a trial, there was a pair of objects on either the right or left side of the stage. Half of the infants were shown the objects with the *same* relation while the other half of the infants were shown the objects with the *different* relation. The objects remained stationary on the stage until the infant accumulated 20 s of looking time at the pair, after which the screen was lowered. The pair was moved to the opposite side of the stage and the familiarization trial was repeated. Across the familiarization trials, infants accumulated 40 s of total looking at the pair.

Test trials. Two types of test trials were presented in sequence to all participants. The first test trial (old) consisted of the pairs from familiarization (fish–fish and crocodile–pig) placed on opposite sides of the stage for 10 s of accumulated infant looking. As in Tyrrell et al. (1991) study, the infants had seen one of these pairs before. The second test trial (novel) consisted of new object pairs (e.g., frog–frog as the *same* relation and dog–dragon as the *different* relation) placed on opposite sides of the stage for 10 s of accumulated infant looking. The amount of time infants fixated on each object pair for each test trial was recorded.

Procedure

During the experiment, the infant sat on the parent's lap in front of the apparatus. The parents were asked to refrain from interacting with their infant during the experiment and to close their eyes during the test trials. All trials ended when the infant accumulated the necessary amount of looking time during each trial (20 s total infant looking for each familiarization trial and 10 s for each test trial). A computer determined the end of the trial and signaled the experimenter to lower the screen. The initial placement (left or right side of the stage) of the familiarization and test stimuli was counterbalanced across infants.

Coding

There was a small hole in the front face of the stage containing a camera that captured a video image of the infant's face. Two research assistants in a separate room viewed this image and coded infants' visual fixations online. Each researcher depressed a computer button when the infant attended to the objects on stage and released the button when the infant looked away. Later, offline frame-by-frame coding to determine left and right looks was conducted using ELAN (2007). Thirty-three percent of infants in each age and condition were recoded by an independent observer. Reliability between observers averaged 92% during familiarization and 91% during test. Three infants (two familiarized to *same* and one familiarized to *different*) looked exclusively to one test stimulus on one of the test trials. Because we were uncertain whether these infants noticed the other test stimulus, infants were given a 2-min break to play in the waiting room and the familiarization and test trials were rerun with a completely different set of objects. We obtained the same results (see below) whether we included these preference scores for these infants or included the original, exclusive test trials.

To measure categorization, a preference score was calculated for each infant by dividing their looking time at the novel-category exemplar during test by their total looking time (10 s).

Results

Preliminary analyses revealed no significant effect of age or sex on the infants' looking times, nor any interactions involving these variables. The subsequent analyses therefore collapsed across these variables. The results, depicted in Figure 1B, reveal that infants familiarized to *same* showed a novelty preference in the old test trial, but no preference in the novel test trial. Infants familiarized to *different* did not differ from chance in either test trial. Overall, infants were not able to abstract the *same–different* relation to the novel pairs, and only those infants familiarized to *same* showed a novelty preference even on the old test trials.

These patterns are borne out by an analysis of variance (ANOVA) on infants' preference scores at test, with condition as a between-subject factor (familiarized to *same* or *different*) and test trial (old or novel) as a within-subject factor. This revealed no main effects, but it did reveal a significant Condition × Test Trial interaction, $F(1, 24) = 5.62$, $p = .026$, $\eta^2 = .19$. Further analyses revealed that within the test trials, infants' performance varied as a function of their familiarization condition, $F(1, 24) = 8.60$, $p = .007$. In the old test trial, infants familiarized to *same* showed a significant novelty preference ($M = 0.70$, $SD = 0.18$), $t(10) = 3.577$, $p = .005$, while infants familiarized to *different* did not significantly differ from chance ($M = 0.43$, $SD = 0.25$), $t(14) = -1.007$, $p = .33$. In contrast, within the novel test trials, performance did not differ between the two familiarization groups, $F(1, 24) = 1.53$, $p = .23$; neither the infants familiarized to *same* ($M = 0.45$, $SD = 0.20$), $t(10) = -0.84$, $p = .42$, nor the infants familiarized to *different* ($M = 0.56$, $SD = 0.24$), $t(14) = 0.96$, $p = .35$, were significantly different from chance.

Discussion

Our attempted replication of the Tyrrell et al. (1991) experiment failed to find evidence that infants encode abstract *same–different* relations from a single training exemplar. That is, there was no generalization to novel test pairs, either for infants familiarized to the *same* pair or for those familiarized to the *different* pair. The only significant finding was that infants familiarized with the *same* pair looked significantly longer at the *different* pair—but only during old test trials, when the *same* pair was identical to the pair seen in familiarization. Infants familiarized to the *different* pair showed chance performance in both test trial types. Thus, we have no grounds to conclude that 7- or 9-month-olds spontaneously encode abstract *same* (or *different*) relations.

The findings by Tyrrell et al. (1991) have been taken as evidence that humans naturally encode abstract relations from very early in development, and may be endowed with a primitive set of abstract relations. Our nonreplication, together with the fact that Tyrell et al.'s studies were inconclusive on this point, calls for a reconsideration of this question. Our findings fail to show support for the hypothesis that infants are endowed with *same–different* relations. At the very least, it appears that these relations are not immediately available to the infant.

Earlier we laid out three possibilities for how analogical ability might develop. The first two both assume that analogical ability is an innate capacity of our species. Possibility 1 adds the assumption that infants are born with a set of primitive relations, including the *same–different* relation. This possibility can be set aside, at least as a solution for how *same* and *different* are acquired. We now turn to the other two possibilities. Possibility 2 is that infants begin life with analogical learning processes that enable them to abstract relations from experience. Possibility 3 is that analogical ability is not an inherent capacity but is formed by combining more basic processes, guided by cultural and linguistic experience.

We now turn to the second possibility: that infants begin life with analogical learning processes that enable them to abstract relations from experience.

Experiment 2 will test this possibility. If we find that 7-month-olds can abstract *same* and *different* relations through analogical abstraction processes, then Possibility 2 will remain a viable explanation. If not, then we will be left with Possibility 3.

Although little is known about the early origins of human analogical ability, there has been considerable research on the development of analogical ability from preschool to adulthood. The available evidence suggests that the same basic analogical mapping processes occur for young children as for older children and adults (Gentner & Medina, 1988, 2003, 2010; Gentner & Namy, 1999; Mix, 2008; Richland, Morrison, & Holyoak, 2006).

Another parallel between children and adults is that the presence of salient objects can interfere with relational mapping. This propensity is especially pronounced early in development. Young children often focus on object matches, whereas older (or more knowledgeable) children focus on relational matches (Gentner & Toupin, 1986; Kotovsky & Gentner, 1996; Paik & Mix, 2006; Richland et al., 2006; Smith, 1984)—a pattern that has been described as the *relational shift* (Gentner, 1988; Gentner & Rattermann, 1991). This shift has been attributed to increases in relational knowledge (Gentner, 2003; Gentner & Rattermann, 1991), to maturational increases in processing capacity (Halford, 1992), and to increases in executive ability, including inhibitory control (Doumas, Hummel, & Sandhofer, 2008; Richland et al., 2006; Thibaut, French, & Vezneva, 2010), and it is possible that all three play a role.

A signature component of analogical learning is that the ability to perceive abstract relational matches can be enhanced by comparing different instances of a relation. For example, Gick and Holyoak (1983) found that comparing two stories that shared causal structure enabled people to generalize that structure and to transfer it to a further situation. These findings are consistent with other research suggesting that the act of comparison entails a structural alignment process that highlights the relational commonalities between the items compared (Gentner & Medina, 1998; Markman & Gentner, 1993). Similar effects of comparison have been found for preschool children (Gentner, Anggoro, & Klibanoff, 2011; Gentner & Namy, 1999; Kotovsky & Gentner, 1996).

The influence of structural alignment in promoting relational focus is a defining characteristic of analogical reasoning in adults (Gentner, Holyoak, & Kokinov, 2001), and the evidence of its influence in children as young as 3 years of age suggests that there may be continuity in analogical learning processes through human development. A second signature component of analogical learning is that attention to individual objects can interfere with relational processing. This is especially true for young children, who show a strong tendency to focus on objects rather than relations. Preschool children perform far worse on relational matching tasks when competing object matches are present (Gentner & Toupin, 1986; Richland et al., 2006), especially if the objects involved are rich and distinctive (DeLoache, 1995; Gentner & Rattermann, 1991; Paik & Mix, 2006). The finding that attention to objects can overshadow attention to relations extends to very young age groups (Casasola, 2005; Maguire, Hirsh-Pasek, Golinkoff, & Brandone, 2008).

A study by Christie and Gentner (2010) illustrates both the appeal of object matches to young children and the effect of comparison in promoting attention to common relational structure. Christie and Gentner taught 3-year-olds names for novel spatial relational patterns, and asked them to choose a further instance of the pattern (see Figure 2). Children given only one standard extended the terms on the basis of object matches rather than common relational structure. For example, given a

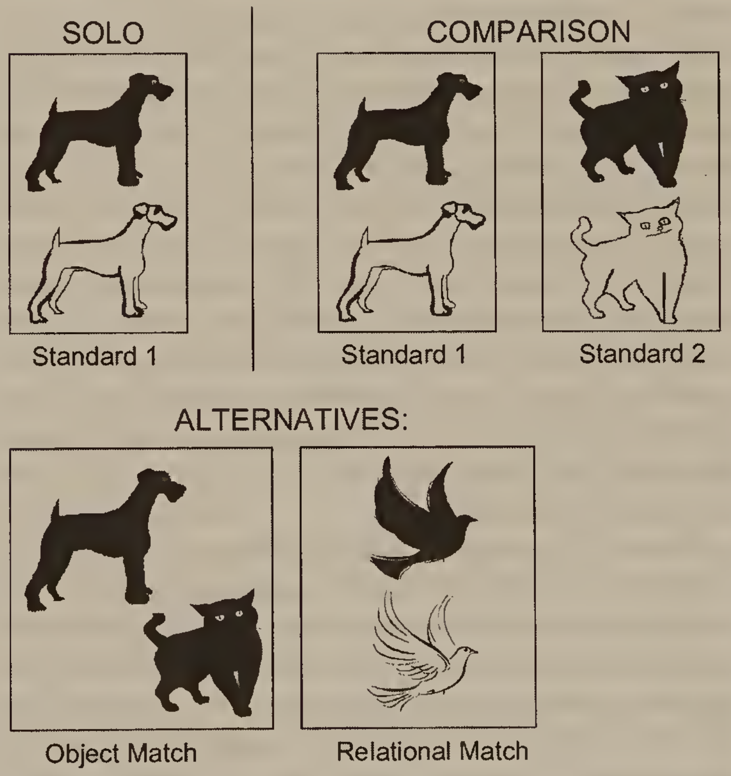


Figure 2. Sample set used in Christie and Gentner (2010) study. Children who saw one standard (either Standard 1 or Standard 2) overwhelmingly chose the object match those who compared the two standards were more likely to choose the relational match.

standard *black dog above white dog; otherwise identical,* the children would choose an alternative showing a black dog and a black cat, rather than one showing the same relational pattern instantiated with different objects. Across two studies, over 90% of the 3-year-olds' responses were to the object match—demonstrating the strength of the object bias in young children. However, when a new group of 3-year-olds was led to compare two standards showing the same relational pattern, the majority of their responses were to the relational match. The effects of comparison were quite striking: The proportion of relational matches rose from 2% to 57%—even though the competing object alternative contained *two* object matches.

Prior Evidence on Early Relational Insight

There is considerable evidence that infants possess some understanding of spatial relations (Casasola, 2005; Casasola & Cohen, 2002; Hespos, Grossman, & Saylor, 2010; Hespos & Piccin, 2009; Hespos & Spelke, 2004; McDonough, Choi, & Mandler, 2003; Quinn, Cummins, Kase, Martin, & Weissman, 1996) and physical relations (Hespos & Baillargeon, 2001, 2006; Wang & Baillargeon, 2008; for a review, see Baillargeon, Li, Gertner, & Wu, 2011). Much of this work has focused on infants' knowledge about interactions between objects. In many of these studies—particularly those using the violation-of-expectation paradigm—the goal is to trace the early development of insight into spatial and physical processes. Although some of this work has investigated the processes involved in gaining knowledge (e.g., Casasola, 2005; Quinn et al., 1996; Wang & Baillargeon, 2008), in most cases the focus has been on discovering what knowledge infants bring to the laboratory. In contrast, our goal is to study infants' learning processes during the experiment itself. We ask whether infants can derive abstract *same–different* relations from specific exemplars by processes of analogical abstraction.

Perhaps the clearest case of relational learning in infants comes from artificial grammar tasks (Aslin & Newport, 2012; Saffran, Aslin, & Newport, 1996), particularly those in which infants learn a relational pattern that can be transferred to new exemplars (Gerken, 2006; Gomez & Gerken, 2000; Johnson et al., 2009; Kirkham, Slemmer, & Johnson, 2002; Marcus, Vijayan, Rao, & Vishton, 1999). For example, in the Marcus et al. (1999) study, when 7.5-month-olds heard 48 examples (16 patterns, three times each) of a syllable pattern such as AAB, they could then discriminate new instances of the AAB pattern from instances of an ABA pattern, even when all the specific syllables were new (see also Gomez & Gerken, 1999). Furthermore, there is evidence that the ability to generalize across such patterns is not restricted to language-like materials; it can operate across a broad range of stimuli, including tones and visual stimuli (Gomez & Gerken, 2000; Johnson et al., 2009; Saffran, Pollak, Seibel, & Shkolnik, 2007).

These experiments focus on learning a particular pattern that emerges during the course of the experiments. How might this be done? Marcus (2000) notes two main approaches—one based on assuming that the child is trying to predict the next syllable, and the other based on assuming a preprocessor that transforms each input item into an abstract form. We suggest a third possibility—namely, that analogical learning mechanisms might take concrete descriptions as input and generalize across a sequence of items to arrive at a more abstract representation. In support of this possibility, Kuehne, Gentner, and Forbus (2000) showed that SEQL (sequential learning engine), a computational model of analogical generalization, can capture the Marcus et al. (1999) findings. SEQL is an offshoot of SME (structure-mapping engine), and operates by using SME to iteratively compare input examples to an ongoing generalization. The model was given the same input as the infants in Marcus et al.: three repetitions of each of the 16 strings, with each syllable encoded as 12 phonemic features (following Elman, 1998). The relational pattern within each string (e.g., AAB) was encoded by Magi, which uses SME to encode symmetry and repetition within an item (Ferguson, 1994). As the strings were presented, SEQL computed a generalization by comparing the first two exemplars (via SME) and storing their common structure, and then incrementally comparing subsequent exemplars to the ongoing abstraction. After three (randomly ordered) repetitions of each example string, SEQL was given two test strings with new syllables. Like the infants in Marcus et al., it was found that the test string with the same relational structure (e.g., CCD) was more similar to its generalization than the one with different structure (e.g., CDC). This rule-like behavior emerged despite the fact that no variables were introduced, and in every run of the simulation, some concrete features were still present at the end. These results provide evidence that analogical generalization could capture the emergence of grammatical patterns. However, because the artificial grammar studies were not designed to test for analogical processes, evidence for (or against) specific signatures of analogical processing is lacking. We return to these issues in the General Discussion.

The Same–Different *Relation in Infants*

Few studies have looked directly at whether infants can detect *same–different* relations. We know from the artificial grammar studies just discussed that infants can discriminate AAB patterns from ABA patterns (Marcus et al., 1999; Saffran et al., 2007), but this does not tell us whether the infants could distinguish AA from AB. That is, the ability to recognize a pattern does not imply the ability to recognize all of its individual subpatterns (Palmer, 1977; Sinha, Balas, Ostrovsky, & Russell, 2006).

More recently, Addyman and Mareschal (2010) showed infants of 4 and 8 months nonrepeating exemplars of either *same* or *different* (up to a maximum of 19 exemplars) during familiarization. Then the infants viewed test trials, presented sequentially. They reported that 8- but not 4-month-olds infants showed evidence of looking longer at the novel than at the familiar relation. It is difficult to interpret these results, because half the infants in both age groups had participated in a prior *same–different* study during the same session. However, it is interesting that although both age groups received over 18 familiarization trials, very few infants showed habituation (a decline in looking time over the course of trials), suggesting that the *same–different* relation did not emerge readily. Still, these results provide encouragement for the idea that by 8 months of age, infants can distinguish *same* and *different* relations under certain conditions. We now turn to the question of how infants extract relational commonalities.

Experiment 2

In Experiment 2, we tested the possibility that infants' relational learning stems from analogical processing—that is, carrying out structural alignment across exemplars and thereby extracting common relational structure. The question then becomes: Can infants derive abstract *same–different* relations from a brief series of exemplars, and if so does this learning process bear the signatures of analogical learning? That is, we ask whether infants engage analogical learning mechanisms during the learning phase to extract relational commonalities from a series of exemplars.

As reviewed above, two key characteristics of analogical learning are: (a) that comparing relationally similar exemplars promotes relational abstraction and (b) that focusing on individual objects interferes with relational processing. In Experiment 2, we ask whether infants would show this learning pattern. That is, we ask, first, whether 7- and 9-month-olds could successfully abstract *same* or *different* relations from multiple exemplars, generalizing the relation to exemplars composed of completely new objects. Such a pattern would suggest that the ability to align sequential exemplars and extract their common relation—a signature of analogical processing in older children—is evident in infants as well. Second, we ask whether infants would also show adverse effects of object focus. If object focus distracts from relational processing, this would constitute a further parallel with analogical learning in older children.

In Experiment 2, infants received training on either *same* or *different* relations. During testing, infants saw pairs instantiating the *same* and *different* relation. The question was whether they would differentiate the familiar relation from the new relation. Critically, each infant saw three kinds of test pairs (as amplified below). Some test pairs were made up of entirely new objects; these tested whether infants had abstracted the relation across the habituation pairs. A second kind of test pairs consisted of objects that had been rendered salient earlier in the session; these tested whether object salience would disrupt relational processing, as in older populations. The third kind of test pairs consisted of objects that had been rendered salient but that had subsequently appeared in the habituation series. These pairs tested whether repeated alignment across pairs would overcome initial object salience, as has also been found for older children. Finally, our method also allowed us to ask whether infants are equally proficient at learning *same* and *different* relations.

The key predictions from the analogical learning account are (a) that alignment across relationally identical pairs should render the common relation more salient and promote abstracting the relation and (b) that attention to objects should disrupt structural alignment and therefore interfere with relational abstraction. Thus, if infants are learning via analogical processes, they should differentiate the familiar relation (e.g., *same,* if habituated to *same*) from the unfamiliar one (e.g., *different*) for novel test pairs.

Method

Participants

The participants were 64 healthy, full-term infants, thirty-two 7-month-olds (17 male and 15 female) ranging from 6 months 9 days to 7 months 22 days (*M* = 6 months 28 days) and thirty-two 9-month-olds (13 male and 18 female) ranging from 8 months 15 days to 9 months 19 days (*M* = 9 months 4 days). Half the infants were assigned to the *same* condition; the other half, to the *different* condition. Seven additional infants were tested but eliminated from the final analysis—three because of fussiness, three because of parental interference, and one due to inattentiveness. The recruitment and socioeconomic status information was similar to Experiment 1.

Apparatus

The stage setup was identical to that used in Experiment 1. However, for this experiment we adopted a habituation–dishabituation paradigm to allow us to examine the rate of learning during training trials. Another change was that the *same* and *different* test trials were presented sequentially, rather than side by side. XHab (Pinto, 1996) was used to record looking times for habituation and test trials during the experiment.

The stimuli consisted of 16 objects, composed of 10 distinctive types (Elmo, blue fuzzy head, pink dotted block, checkered cylinder, dragon, red striped block, camel, pig, green puff balls on wood, and red/green foam block), with each pair placed on a 26.5 × 15.5 cm piece of cardboard covered with contact paper (see Figure 3). Twelve of the objects (composed of seven distinct types) were presented in the playroom before the experiment began. Of these, eight (four pairs) were also used in the habituation phase and in the test phase, and three were used in the test phase only. Four pairs of objects were used in habituation and six were used in the test.

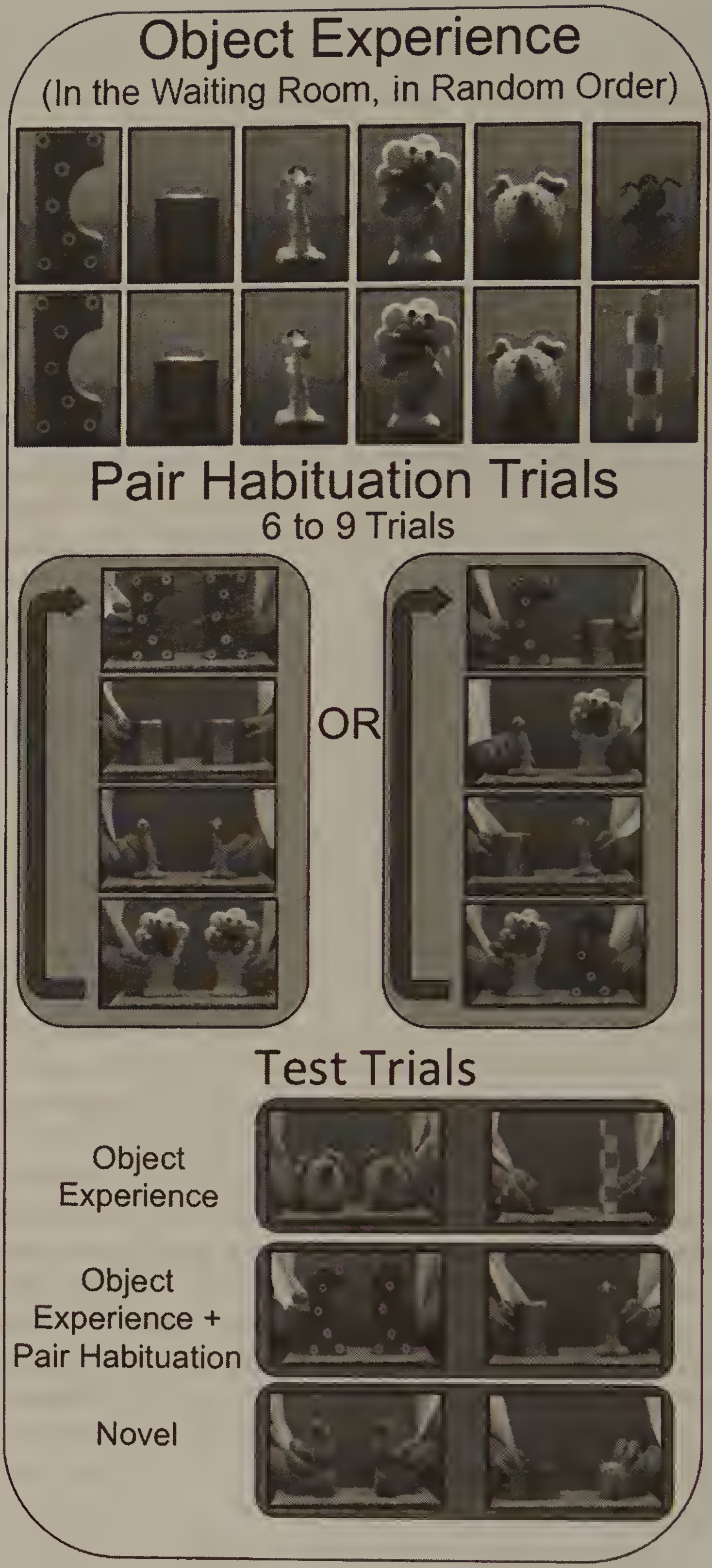


Figure 3. Schematic of events in Experiment 2. In the waiting room infants saw a subset of the individual toys before the experiment. Infants were habituated to four pairs of objects, either same or different. In six sequential test trials looking time was recorded to the novel and familiar relational pairs in three different types of test trials (object experience, object experience + pair habituation, and novel).

Events

Object experience phase. In the object experience phase, we manipulated object focus by showing some of the test objects to the infants in the playroom prior to the experiments, during naturalistic play interactions. We held up the objects and made comments such as, "Look!" and "See this one?" The objects were presented individually and in no particular order for approximately 5 s for each object. The objects consisted of all of the objects used in the habituation trials (two of each Elmos, camels, red blocks, and pink blocks), as well as a set of objects used only in the object experience test trials (two blue fuzzy heads, one checkered cylinder, and one dragon).

Pair habituation trials. To engage infants' attention, in both habituation and test trials the pairs were moved during the trial. When the screen went up at the start of a trial, a pair of objects rested on the cardboard plate on the stage. Each of the experimenter's hands grasped one of the objects in the pair and raised the objects straight up (1 s), tilted the objects to the left (1 s), returned them to the center (1 s), tilted them to right (1 s), returned them to the center (1 s), returned them to the plate (1 s), and paused on the plate (2 s). This 8-s cycle repeated continuously until the trial ended. In the *same* condition, infants saw habituation trials in which the pairs of objects were the *same* (see Figure 3). In the *different* condition, infants saw habituation trials in which the pairs of objects were *different*. Infants saw four sets of object pairs during habituation trials. These four pairs were repeated until the infant reached the habituation criterion. The order of presentation was counterbalanced across infants.

Test trials. Infants received six test trials. In each test trial, infants viewed one pair of objects, presented in the same manner as in the habituation trials, while their looking time was recorded. Each infant received test trials with both *same* and *different* pairs of objects, presented in alternation, with order counterbalanced across infants. There were three kinds of test trials: (A) pairs consisting of objects that the infant had only experienced individually in the playroom (object experience), (B) pairs consisting of objects that had been seen individually in the playroom and then in pairs during habituation (object experience + pair habituation), and (C) pairs consisting of objects that were completely novel, in that the infants saw them for the first time during the test trials (novel). There were three trial orders in which the test items could appear: ABC, BCA, or CAB. This allowed each type of test pair to be first for a third of the infants. This ensured that the order of the test pairs was counterbalanced across infants and that any effects of test trial type were distinguishable from effects due to the order of the test pairs.

Procedure

During the experiment, the infant sat on the parent's lap in front of the apparatus. The parents were asked to refrain from interacting with their infant during the experiment and to close their eyes during the test trials. All trials ended when the infant either looked away for 2 consecutive seconds after having looked at the event for at least 2 s, or looked at the event for 60 cumulative seconds without looking away for 2 consecutive seconds. A computer determined the end of the trial and signaled the experimenter to lower the screen. The habituation criterion was at least a 50% decline in total looking duration from the first three to the last three habituation trials or a maximum of nine trials. Each infant viewed six test trials, alternating between *same* or *different* pairs of objects. The type of test event shown first and the order of the test trials was counterbalanced across infants. Interobserver agreement was measured for all infants and averaged 94% per trial per infant.

Results

Preliminary analyses revealed no significant effect of gender or test trial order on the infants' looking times; the subsequent analyses therefore collapsed across these variables.

Overall

The first analyses addressed our central question: whether infants are capable of abstracting *same* and *different* relations from multiple exemplars. The answer appears to be yes; infants looked significantly longer at the novel relation than at the familiar relation during the test trials for both *same* and *different* (see Figure 4). Collapsing across all test trial types, 49 of the 64 infants (77%) had longer looking times for the novel test events than for the familiar test events ($p < .001$, binomial comparison). Across all participants, the average looking times were 18.32 s ($SD = 8.34$) for the novel relations and 14.50 s ($SD = 7.05$) for the familiar relations, $F(1, 63) = 19.87$, $p < .001$, $\eta^2 = .24$. Overall, infants detected a change from the familiar to the novel relation.

Effects of Prior Experience

The next set of analyses tested whether prior experience in which attention was called to individual objects would impede relational learning for pairs containing these objects; and if so, whether this adverse effect would be mitigated by exposure to pairs containing these objects during habituation. In both cases, the answer appears to be yes. An ANOVA comparing the within-subject factors of

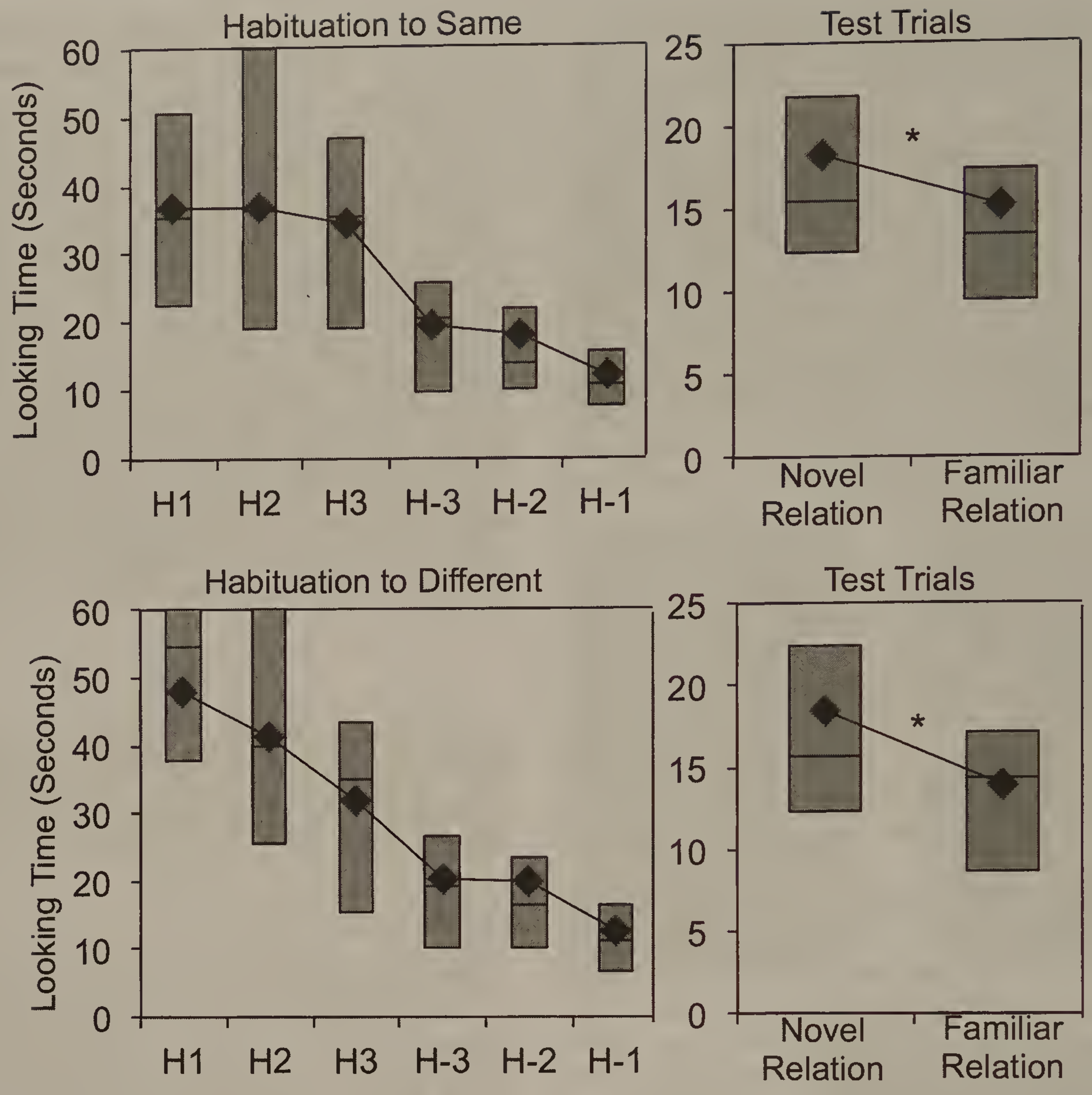


Figure 4. Box plots showing mean looking times during the habituation and test trials in Experiment 2. At test, preference for the novel relation was significant in both *same* and *different* conditions. The habituation graphs include results for the first three trials (H1, H2, H3) and the last three trials before the habituation criterion was met (H-3, H-2, H-1). Black diamonds represent means, the central line in each box is the median, and the upper and lower portions of each box represent the third and first quartiles respectively.
*Significant difference between novel and familiar test trials.

relation (novel, familiar), and prior experience (object experience, object experience + pair habituation, novel) revealed a significant effect of relation, $F(1, 63) = 20.01$, $p < .001$, $\eta^2 = .24$, and a Relation × Prior Experience interaction, $F(2, 62) = 3.17$, $p = .049$, $\eta^2 = .093$ (see Figure 5A). Further analyses revealed that infants looked significantly longer to the novel relation even with completely new objects (novel test trials), $F(1, 63) = 14.15$, $p < .001$, $\eta^2 = .18$, indicating that they successfully generalized the abstract relation to new objects presented for the first time in test trials. In the trials in which attention was called to the individual objects (object experience test trials), the generalization of the relation was impeded; infants showed no significant difference in looking to the novel and familiar relations, $F(1, 63) = 0.114$, $p = .74$, $\eta^2 = .002$. However, this adverse effect of object focus was mitigated by exposure to pairs containing these objects during habituation; on these trials, infants looked significantly longer to the novel relation (object experience + pair habituation test trials), $F(1, 63) = 8.70$, $p = .004$, $\eta^2 = .12$.

To isolate the effect of prior experience and training we looked at the recovery scores of the first pair of tests (the first novel relation test trial and the first familiar relation test trial). Recovery scores are a standard looking-time analysis for habituation–dishabituation paradigms, involving three comparisons. The first is a *t* test comparing the last habituation to the first novel test trial. If infants perceive the test event as novel, then there should be a significant increase in looking time from the last habituation trial to the novel test trial. The sec-

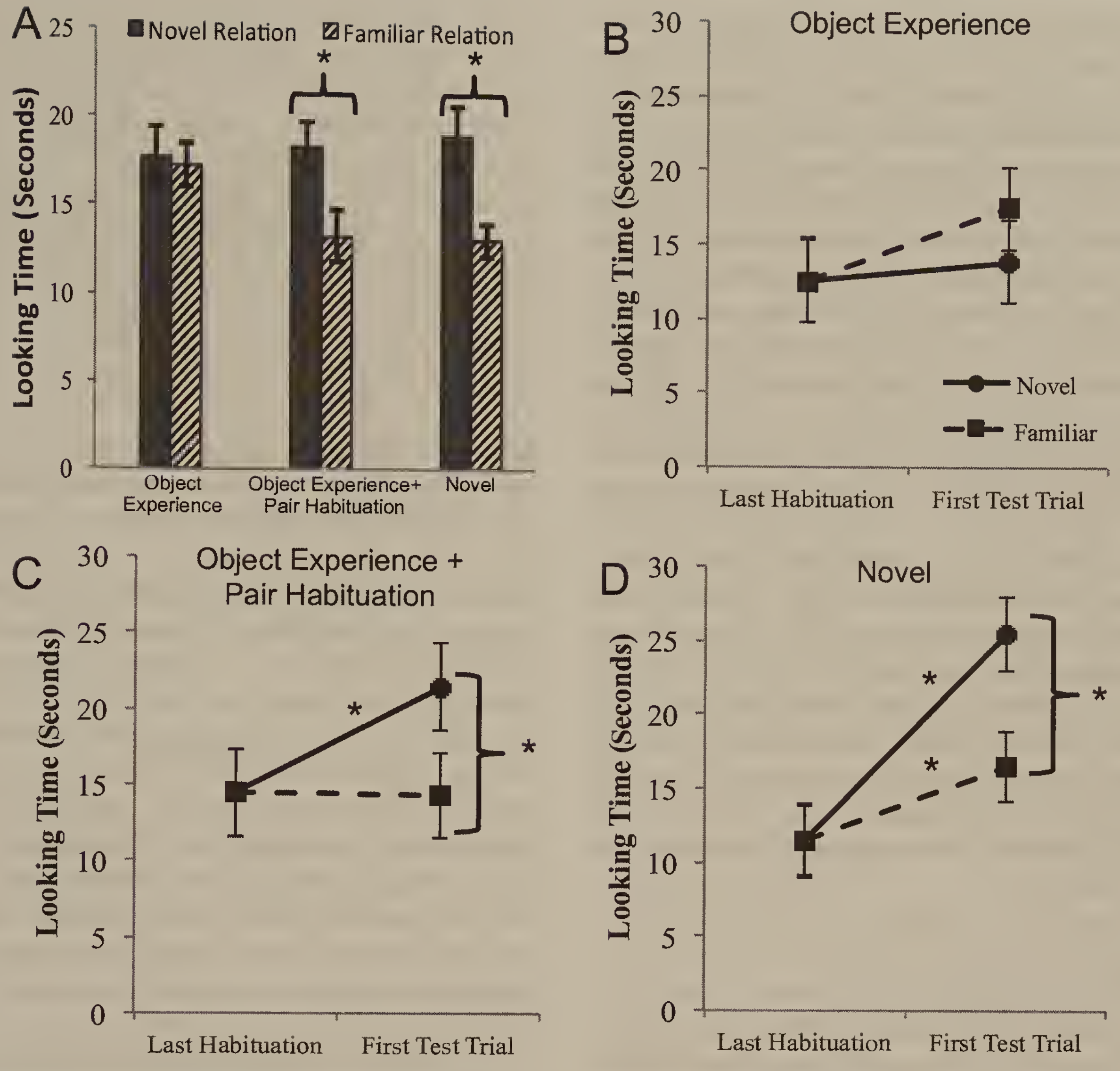


Figure 5. Looking time in Experiment 2. (A) The overall looking time to novel and familiar relations in the three types of test trials. The change in looking time from the last habituation to the first test pair (solid line for the novel event, dotted line for the familiar event) for (B) object experience, (C) object experience + pair habituation, and (D) novel test trials.
$^*p < .05$.

ond is a *t* test comparing the last habituation to the first familiar test trial. If infants generalize their habituated response to the familiar test event, then looking time should not be statistically different between the last habituation trial and the familiar test trial. For completeness, the third analysis is a *t* test comparing the novel and familiar test trials, which asks whether infants overall look significantly longer at the novel than at the familiar test trials. The recovery analysis goes beyond the significant interaction of relation by prior experience above, in that it focuses on the first test pair, removing the influence of prior test trials on the results. As we found no effect of condition (habituation to *same* or *different*) or age group (7- or 9-month-olds) in the overall analysis of the prior experience effects, we collapse across these in the recovery scores analysis.

For infants who saw object experience as the first test trials, there was no change in looking time from the last habituation trial ($M = 12.4$ s, $SD = 9.5$ s) to the novel relation ($M = 15.4$ s, $SD = 11.7$ s), $t(22) = 1.00$, $p = .33$, or from the last habituation trial to the familiar relation ($M = 17.1$ s, $SD = 10.8$ s), $t(22) = 1.88$, $p = .08$. In addition, there was no overall difference between novel-relation and familiar-relation test trials, $t(22) = 0.722$, $p = .478$ (see Figure 5B). This pattern suggests that when the first test trial consisted of high-salience objects, infants did not detect the *same–different* relation on that trial. In contrast, infants who saw object experience + pair habituation as their first test trials showed evidence of having encoded the relation. That is, they showed a significant increase from the last habituation trial ($M = 14.71$ s, $SD = 7.0$ s) to the novel-relation test trial

($M = 22.2$ s, $SD = 13.0$ s), $t(20) = 2.45$, $p = .024$, but not to the familiar-relation test trial ($M = 14.6$ s, $SD = 13.3$ s), $t(20) = 0.49$, $p = .96$. Furthermore, there was a significant overall difference between the novel-relation and the familiar-relation test trials, $t(20) = 2.16$, $p = .043$ (see Figure 5C). This is the classic dishabituation pattern. The infants dishabituated to the novel-relation event and generalized to the familiar-relation event. Infants who saw novel as their first test trials also showed significant recovery from the last habituation ($M = 11.5$ s, $SD = 7.4$ s) to the novel relation ($M = 25.5$ s, $SD = 17.9$ s), $t(19) = 3.17$, $p = .005$. In addition, they showed significant recovery to the familiar relation ($M = 16.5$ s, $SD = 7.5$ s), $t(19) = 2.48$, $p = .023$. Importantly, there was a significant overall difference between the novel-relation and familiar-relation test trials, $t(19) = 2.38$, $p = .028$. That infants dishabituated to both the novel-relation and familiar-relation test events suggests that they detected the objects as novel. However, the significant difference between the novel-relation and familiar-relation test events indicates that they also discriminated between the novel and familiar relations (see Figure 5D).

Effects of Age

To investigate whether there was an effect of age, we did an ANOVA with the within-subject variables of relation (novel vs. familiar) and test trial (first, second, or third pair) and the between-subject factor of age (7 or 9 months). This revealed a marginal three-way interaction, $F(2, 61) = 2.78$, $p = .07$, $\eta^2 = .08$. To make the analysis easier to perceive, we translated the relation factor into a difference score by subtracting the looking times at the familiar from the novel events for each test trial pair. These data are represented in Figure 6. They demonstrate that 9-month-old infants show a large difference in looking time (looking longer at the novel relation pair) on the first test pair, which diminishes over the course of test trials. In contrast, the 7-month-old infants show an opposite time course: They also look longer at the novel relation, but their difference in looking time increases across the test sets.

Effects of Relation

Our next question was whether infants would find it easier to align the *same* relation than the *different* relation. Perhaps surprisingly, we found no difference between these two conditions. As

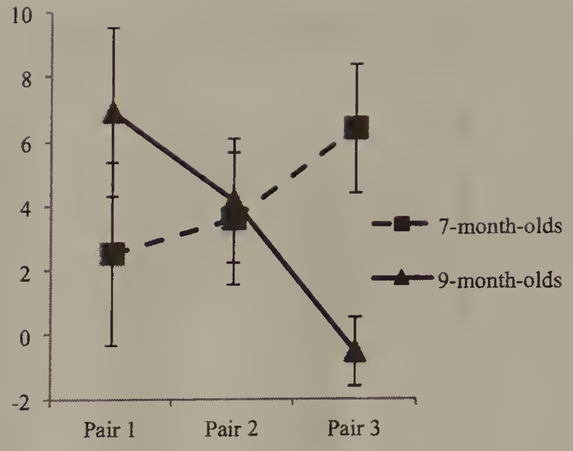


Figure 6. A line graph showing the difference score (looking time to the novel relation − looking time to the familiar relation) across the three pairs of test events and separated by age.

shown in Figure 3, infants detected the novel relation at test regardless of whether they were habituated to *same* or to *different*. Further analyses revealed that there was no difference in number of trials to meet the habituation criterion for *same* versus *different* ($M = 7.42$ and 7.24, respectively, range = 6–9), $F(1, 62) < 1$, $p = .57$, or in the amount of looking time that infants accumulated during habituation ($M = 202$ s, $SD = 75$ s; $M = 218$ s, $SD = 78$ s, respectively), $F(1, 62) < 1$, $p = .38$. Overall, we found no evidence that the *same* relation was easier to learn than the *different* relation.

This pattern of results was consistent and significant within each habituation condition. For the habituation to *same* condition, 23 of 31 infants (74%) looked longer at the *different* test trials ($p = .011$, binomial comparison), and an ANOVA showed a significant main effect of test relation, $F(1, 30) = 5.01$, $p = .03$, $\eta^2 = .14$. For the habituation to *different* condition, 26 of 33 (79%) infants looked longer at the *same* test trials ($p = .001$, binomial comparison), and the ANOVA again showed a main effect of test relation, $F(1, 32) = 18.27$, $p < .001$, $\eta^2 = .36$. We also found no reliable differences between *same* and *different* test trials when these relations were both novel, $F(2, 61) < 1$, or when they were both familiar, $F(2, 61) < 1$. Thus, we found no evidence in this study that *same* is easier to learn than *different*.

Effects of Habituation Phase

Thirty-eight percent of the infants met the habituation criterion in the minimum of six trials, 54% met the criterion in seven to nine trials, and 16% of infants did not meet the habituation criteria after the maximum of nine trials. On average, infants reached the habituation criterion after seven trials. Thus, infants typically saw three of the four training pairs a second time. An ANOVA of the accumulated looking during habituation reveal no main effect of age group, $F(1, 60) < 1$; no main effect of condition, $F(1, 60) = 1.13$, $p > .05$; and no interactions, $F(1, 60) < 1$. Infants accumulated an average of 210.6 s ($SD = 76.7$ s) of looking during the habituation phase. Additionally, there was no correlation between the duration of looking time during habituation trials and preference for the novel relation, $r(62) = -0.02$. A median split separating the shortest and the longest total looking times during habituation revealed that both groups looked significantly longer at the novel *same–different* relation than to the familiar relation (both $Fs > 9.6$, both $ps < .005$). Thus, the amount of looking time during habituation had no impact on performance during test trials.

Discussion

The first question in this study was whether infants were capable of detecting and generalizing an abstract *same–different* relation. The results indicate that the answer is yes. Infants were able to generalize their learning to purely novel test pairs, showing that they had formed an abstract relation. Furthermore, infants habituated after just seven trials (on average) consisting of four distinct pairs plus repetitions, suggesting that the abstraction process was quite efficient. We infer that comparing across exemplars of the relation allowed the infant to abstract the common relational pattern—*same* or *different*. This pattern—comparing exemplars of a relation renders the relation more salient—is well established in studies with both adults (Gick & Holyoak, 1983) and children (Childers & Paik, 2009; Christie & Gentner, 2010; Gentner et al., 2011; Kotovsky & Gentner, 1996; Maguire et al., 2008). These findings suggest that this signature of analogical reasoning is present as early as 7 months of age.

We note that there is a way to save Possibility 1 here (that infants are born with abstract representations of *same* and *different*). It could be that the infants in Experiment 2 generalized better than those in Experiment 1 simply because they had more opportunities to encode the relation (and more time to encode overall) than did those in Experiment 1. Although we cannot rule out this possibility, the overall fit with predictions suggests that analogical generalization is the more parsimonious explanation.

The second question was whether there would be evidence of interference from attention to individual objects. Previous research on analogical learning has found that salient objects impede the structural alignment process (Gentner & Rattermann, 1991; Gentner & Toupin, 1986; Richland et al., 2006). To address this question, we systematically manipulated the infants' prior experience with the objects. Prior experience with the individual objects (object experience) interfered with the infants' ability to detect the *same–different* relation. However, this detrimental effect of prior object experience was overcome if infants then experienced those *same* objects within pairs during habituation. In that case, the infants successfully distinguished novel from familiar relations on the test pairs (object experience + pair habituation). Since infants also made this distinction for novel test pairs (consisting of objects they had not been seen before), their failure to show an effect for object experience test pairs is strong evidence that object focus interfered with their ability to detect the relation—consistent with findings demonstrating the detrimental effects of object focus on analogical processing.

Finally, we discovered an intriguing age effect. The older infants showed a large difference in looking time on the first test pair; then the difference attenuated over the subsequent pairs. The younger infants showed the opposite pattern. There was only a small difference in looking across the first two test trial types, but the difference increased across test trials. These differences suggest that (as one might have expected) the relational abstraction process had progressed further during habituation in the older infants than in the younger infants.

Our design also allowed us to ask whether performance would be better with the *same* relation than with the *different* relation. Intuitively, it seems possible that *same* is an elemental relation, while *different* may be represented as "not *same*" (Clark & Chase, 1972; Hochmann, Mody, & Carey, 2013), suggesting that children should more readily be able to abstract the *same* relation from experience than the *different* relation. However, prior evidence is mixed on this point (Addyman & Mareschal, 2010; Smith, Redford, Haas, Coutinho, & Couchman, 2008). In our study, we found no measurable difference in the likelihood of learning a *same* versus a *different* relation. An intriguing question is whether infants younger than 7 months would show an advantage for *same* over *different*. If so, then the idea that *different* is represented as *not same* might still be maintained.

What was learned? After seeing a series of *same* or *different* exemplars in Experiment 2, infants represented an abstract relation of *same* or *different* that was not evident when they saw a single example in Experiment 1. How might this have happened? One possibility is that the infants had absolutely no notion of *same* and *different* relations before experiencing this series in Experiment 2. A more plausible explanation is that the infants were able to encode some initial form of the relation, at least for some of the pairs, but that these initial representations were highly context specific. In this case, the effect of structural alignment across examples was to lead the infants to rerepresent the relations in a more abstract way. On this account, infants would fail to apply the relation to a new pair after just one initial example (as we found in Experiment 1), but would be able to do so after aligning across examples (as we found in Experiment 2).

There is considerable evidence for this kind of sequential analogical abstraction (e.g., Gentner, Loewenstein & Hung, 2007; Gentner et al., 2011; Haryu, Imai, & Okada, 2011; Kotovsky & Gentner, 1996). For example, in the Kotovsky and Gentner (1996) study, 4-year-old children were initially able to accurately match higher order spatial relational patterns (such as monotonic increase from left to right) within dimension but not across dimension. For example, they could match *blue circles increasing in size* with *red squares increasing in size*, but not with *red squares increasing in darkness from left to right*. But after experiencing several within-dimension trials (with no feedback), the children were able to perceive the cross-dimensional match. Gentner, Rattermann, Markman, and Kotovsky (1995) modeled these results by assuming that children initially represented the relations in a dimensionally specific manner (roughly "bigger than" and "darker than") and that structural alignment across examples prompted them to rerepresent the relations in a more abstract way (roughly "greater along a dimension than"). We think something like this could apply in the current study.

General Discussion

Analogical ability—the ability to perceive and reason about common relational patterns across disparate situations—is central to human intelligence. The available evidence indicates that humans are far better at representing and reasoning about relations than are other species. Recent theoretical accounts have argued that this difference in relational ability is the central cognitive difference between humans and other primates (Gentner, 2003, 2010; Penn et al., 2008). The question of how human analogical ability develops is therefore important not only to theories of human development but also to accounts of comparative cognition and of how to characterize human intelligence.

In this article, we asked whether analogical ability is present in human infants, using the simplest and most basic relation—that of sameness or difference between two things. We proposed three broad possibilities: (a) that humans are born with analogical ability and a set of primitive relations for encoding early experiences, (b) that humans are born with analogical ability but not with a set of primitive relations, and (c) that analogical ability is not inherent in humans but emerges from linguistic and cultural experiences. We began by revisiting Tyrrell et al. (1991) influential study, which suggested that infants—even at 7 months of age—form an abstract *same* relation from just one exemplar of a *same* pair, and likewise for *different*. If indeed infants encode an abstract relation from a single exemplar, this would suggest that they already have a basic set of abstract relations, which can be applied without prior experience. These findings suggested a qualitative difference between humans and other primates, and have been widely cited in the primate literature. Because of the importance of this question, we first tried to replicate this finding.

Our data do not support this conclusion. In our attempted replication of the Tyrrell et al. (1991) study (Experiment 1), we found no evidence that prelinguistic human infants spontaneously encode abstract *same–different* matches from a single exemplar. The infants in our replication study showed a novelty response only when the familiar-relation pair was precisely identical (in objects as well as relation) to one that they had experienced before; there is no indication that they went beyond recognizing the identical item they had seen before. They failed to distinguish new *same* pairs from new *different* pairs, as would have been required to conclude that they already have represented an abstract *same–different* relation.

Our second question, given that 7- and 9-month-old infants do not immediately encode abstract *same* and *different* relations, was whether they could form these abstract relations via analogical learning processes. To address this, we asked whether the signatures of analogical learning seen in older children—the positive effects of multiple exemplars and the adverse effects of object focus—would be evident in infants' relational learning. We found that 7-month-old infants can indeed abstract a *same* or *different* relation across multiple exemplars. Although one exemplar was not sufficient (Experiment 1), the results of Experiment 2 showed that if infants are given a series of exemplars, they can form a *same* or *different* relation; that is, they can distinguish between *same* and *different* pairs even when these pairs are made up of totally novel objects. Only a modest number of exemplars was required for this learning. Infants saw four exemplars of *same* (or *different*) pairs, with two or three of them repeated, for a total of six to nine exemplars. Although we made no specific predictions here, this finding is consistent with the typical finding with older children, that comparing even a few exemplars of a relation can lead to significant learning.

Experiment 2 also bore out the second signature of analogical learning—namely, that object focus interferes with relational matching. When infants saw pairs consisting of objects that had been demonstrated as single objects in the playroom prior to the study, they failed to show a distinction between *same* and *different* pairs unless they had subsequently seen the objects in pairs during habituation. This is consistent with other findings demonstrating that a focus on individual objects can disrupt relational processing (Casasola, 2005; Christie & Gentner, 2010; Gentner & Toupin, 1986; Maguire et al., 2008; Richland et al., 2006). It appears that there is continuity in the signatures of relational learning in the earliest evidence of this ability.

Our findings support the broad position that humans are born with analogical skills and use these skills to extract relational information from experiences. Thus, they argue against the third possibility suggested earlier—that analogical ability emerges developmentally as an amalgam of other cognitive processes, rather than being an inherent cognitive capacity. Of course, this is not to suggest that this ability is fully developed at an early age. Clearly, there are developmental changes and influences from language and cultural experiences. Nevertheless these are the first data that demonstrate that signatures of the analogical processing are evident early in development.

Broader Implications and Future Directions

Returning to the question of how to characterize human relational ability, we found no support for the strong view that humans are endowed with a set of abstract relational representations. Rather, our findings suggest that humans are born with considerable analogical processing ability, with which they can detect abstract relations. This characterization is consistent with there being a continuum of analogical ability between our species and other apes, rather than a sharp dichotomy. As we see it, the evidence from primate studies suggests that other apes possess some analogical reasoning ability, though to a substantially lesser degree than humans (Christie, Gentner, Call, & Haun, 2015; Haun & Call, 2009; Premack, 1983; Thompson et al., 1997). Indeed, recent studies have suggested that even more distant relatives (hooded crows) have some analogical ability and can pass the relational-match-to-sample test under some circumstances (Smirnova, Zorina, Obozova, & Wasserman, 2015).

Our findings also bear on the role of language in analogical learning. Gentner and colleagues have suggested that acquiring and using relational language is instrumental in the development of analogical processing (Christie & Gentner, 2014; Gentner, 2010; Gentner & Rattermann, 1991; Gentner et al., 2011). Our results clearly place limits on these claims. The infants in our study were able to carry out analogical generalization of a *same–different* relation prior to learning terms for *same* and *different*. We can infer that analogical ability does not *depend* on the possession of relational language. However, research with older children shows that language can augment our relational capacity, by inviting comparison (Namy & Gentner, 2002) and by heightening attention to particular relations (Casasola, 2005; Gentner, Özyürek, Gürcanli, & Goldin-Meadow, 2013). Thus, it is an open question whether we might see language effects later in development. We know that relational terms are acquired later than object-reference terms (Bornstein et al., 2004; Gentner, 1982; Gentner & Boroditsky, 2009); indeed, infants do not appear to understand relational words until the end of the 1st year (Bergelson & Swingley, 2013). Thus, future research should investigate whether and when we might see language effects for *same–different* processing in infants.

Earlier we raised the question of whether infants' abilities in artificial grammar tasks could be related to the analogical learning processes studied here. We think it quite possible that some of the same processes are involved. Artificial grammar tasks capture two crucial requirements for grammar learning: first, the ability to learn sequential transition probabilities (Saffran et al., 1996), and second, the ability to extract and learn parallel structure across a set of exemplars (Marcus, 2000; see also Gomez & Gerken, 2000). It is this second ability, often termed *rule learning*, that is most closely connected with analogical generalization. That infants can learn this kind of general structure was demonstrated by Marcus et al. (1999) in their seminal study showing that 7-month-olds exposed to a series of exemplars such as *pa ti pa* and *li na li* formed an abstract ABA structure—as evidenced by their showing a novelty response to a new pattern (e.g., *wo wo fe*) but not to an old pattern (*wo fe wo*).

Is this kind of relational generalization a product of analogical learning? There is some suggestive evidence for this speculation. First, the same kind of generalization patterns have been found for visual stimuli (Johnson et al., 2009; Saffran et al., 2007). This is consistent with the operation of a domain-general learning process such as analogy, rather than a dedicated language-specific process. Second, studies by Gerken (2006) show that when the exemplars all contain the same extra element, the infants' generalization retains that element—consistent with analogical learning patterns, which typically show conservative generalization, especially early in learning. A third line of evidence, also noted above, is that we have been able to simulate the Marcus et al. (1999) study with a computational simulation of analogical generalization (Kuehne et al., 2000). Unlike most attempts to simulate these results, the analogical generalization mechanism requires only the same 48 exemplars given to the infants, and can succeed even with noise added to the data. Of course, these are only plausibility arguments. An important future direction will be to test directly whether analogical generalization processes operate in artificial grammar learning.

Another open question is how to square our claims concerning the continuity of relational learning processes with the fact that even 2½-year-old children have difficulty with *same–different* relational matching (Christie & Gentner, 2014). One possible explanation lies in the difference between the tasks typically given to older and younger children. For example, there is evidence that 18-month-old children can succeed in abstracting *same–different* relations across examples in service of a causal learning task (Walker & Gopnik, 2014).

Our findings also raise questions for future infant research. First, what is the earlier course of relational learning? Would we see analogical learning even among much younger infants, and if so, would they require more extensive training (e.g., more exemplars, more repetitions or longer exposure times) to abstract relations? Research with newborn infants by Gervain, Berent, and Werker (2012) suggests that some sensitivity to relational patterns may appear very early. Using the functional near-infrared spectroscopy technique, they found that newborns formed a different neural response to AAB patterns than to ABB patterns after exposure to blocks of syllables in the two patterns. Whether and when this ability would allow infants to generalize to distinguish new instances of one pattern from another is an intriguing question.

In conclusion, our findings show that analogical ability is present in infants by at least 7 months of age. Our results also suggest that there is continuity in the analogical learning process across development, and leave open the possibility of continuity across species.

References

Addyman, C., & Mareschal, D. (2010). The perceptual origins of the abstract same/different concept in human infants. *Animal Cognition, 13,* 817–833. doi:10.1007/s10071-010-0330-0

Aslin, R. N., & Newport, E. L. (2012). Statistical learning: From acquiring specific items to forming general rules. *Current Directions in Psychological Science, 21,* 170–176. doi:10.1177/0963721412436806

Baillargeon, R., Li, J., Gertner, Y., & Wu, D. (2011). How do infants reason about physical events? In U. Goswami (Ed.), *The Wiley-Blackwell handbook of childhood cognitive development* (2nd ed., pp. 11–48). Oxford, UK: Blackwell. doi:10.1002/9781444325485.ch1

Bergelson, E., & Swingley, D. (2013). The acquisition of abstract words by young infants. *Cognition, 127,* 391–397. doi:10.1016/j.cognition.2013.02.011

Bornstein, M. H., Cote, L. R., Maital, S., Painter, K., Park, S.-Y., Pascual, L., . . . Vyt, A. (2004). Crosslinguistic analysis of vocabulary in young children: Spanish, Dutch, French, Hebrew, Italian, Korean, and American English. *Child Development, 75,* 1115–1139. doi:10.1111/j.1467-8624.2004.00729.x

Casasola, M. (2005). When less is more: How infants learn to form an abstract categorical representation of support. *Child Development, 76,* 279–290. doi:10.1111/j.1467-8624.2005.00844.x

Casasola, M., & Cohen, L. B. (2002). Infant categorization of containment, support, and tight-fit spatial relationships. *Developmental Science, 5,* 247–264. doi:10.1111/1467-7687.00226

Childers, J. B., & Paik, J. H. (2009). Korean- and English-speaking children use cross-situational information to learn novel predicate terms. *Journal of Child Language, 36,* 201–224. doi:10.1017/s0305000908008891

Christie, S., & Gentner, D. (2010). Where hypotheses come from: Learning new relations by structural alignment. *Journal of Cognition and Development, 11,* 356–373 doi:10.1080/15248371003700015

Christie, S., & Gentner, D. (2014). Language helps children succeed on a classic analogy task. *Cognitive Science, 38*, 383–397. doi:10.1111/cogs.12099

Christie, S., Gentner, D., Call, J., & Haun, D. (2015). *Sensitivity to relational similarity and object similarity in apes and children.*

Clark, E. V. (1973). What's in a word? On the child's acquisition of semantics in his first language. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 65–110). New York, NY: Academic Press.

Clark, H. H., & Chase, W. G. (1972). On the process of comparing sentences against pictures. *Cognitive Psychology, 3*, 472–517. doi:10.1016/0010-0285(72)90019-9

DeLoache, J. S. (1995). Early symbol understanding and use. In D. L. Medin (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 33, pp. 65–114). San Diego, CA: Academic Press.

Dewar, K. M., & Xu, F. (2010). Induction, overhypothesis, and the origin of abstract knowledge: Evidence from 9-month-old infants. *Psychological Science, 21*, 1871–1877. doi:10.1093/acprof:oso/9780199216895.003.0011

Doumas, L. A. A., Hummel, J. E., & Sandhofer, C. M. (2008). A theory of the discovery and predication of relational concepts. *Psychological Review, 115*, 1–43. doi:10.1037/0033-295x.115.1.1

ELAN. (2007). *ELAN*. Nijmegen, Netherlands: Language Archiving Technologies.

Elman, J. (1998). *Generalization, simple recurrent networks, and the emergence of structure.* Presented at the annual conference of the Cognitive Science Society, Madison, WI.

Fagot, J., & Thompson, R. K. R. (2011). Generalized relational matching by guinea baboons (Papio papio) in two-by-two-item analogy problems. *Psychological Science, 22*, 1304–1309. doi:10.1177/0956797611422916

Ferguson, R. W. (1994). MAGI: Analogy-based encoding using regularity and symmetry. In A. Ram & K. Eiselt (Eds.), *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society* (pp. 283–288). New York, NY: Psychology Press.

Flemming, T. M., Beran, M. J., & Washburn, D. A. (2007). Disconnect in concept learning by rhesus monkeys (*Macaca mulatta*): Judgment of relations and relations-between-relations. *Journal of Experimental Psychology: Animal Behavior Processes, 33*, 55–63. doi:10.1037/0097-7403.33.1.55

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. A. Kuczaj (Ed.), *Language development: Language thought and culture* (Vol. 2, pp. 301–334). Hillsdale, NJ: Erlbaum.

Gentner, D. (1988). Metaphor as structure mapping: The relational shift. *Child Development, 59*, 47–59. doi:10.2307/11303

Gentner, D. (2003). Why we're so smart. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 195–235). Cambridge, MA: MIT Press.

Gentner, D. (2010). Bootstrapping the mind: Analogical processes and symbol systems. *Cognitive Science, 34*, 752–775. doi:10.1111/j.1551-6709.2010.01114.x

Gentner, D., Anggoro, F. K., & Klibanoff, R. S. (2011). Structure mapping and relational language support children's learning of relational categories. *Child Development, 82*, 1173–1188. doi:10.1111/j.1467-8624.2011.01599.x

Gentner, D., & Boroditsky, L. (2009). Early acquisition of nouns and verbs: Evidence from Navajo. In V. Gathercole (Ed.), *Routes to language: Studies in honor of Melissa Bowerman* (pp. 5–36). New York, NY: Psychology Press.

Gentner, D., Holyoak, K. J., & Kokinov, B. N. (2001). *The analogical mind: Perspectives from cognitive science.* Cambridge, MA: MIT Press.

Gentner, D., Loewenstein, J., & Hung, B. (2007). Comparison facilitates children's learning of names for parts. *Journal of Cognition and Development, 8*, 285–307.

Gentner, D., & Medina, J. (1998). Similarity and the development of rules. *Cognition, 65*, 263–297. doi:10.1016/s0010-0277(98)00002-x

Gentner, D., & Namy, L. (1999). Comparison in the development of categories. *Cognitive Development, 14*, 487–513. doi:10.1016/s0885-2014(99)00016-7

Gentner, D., Özyürek, A., Gürcanli, Ö., & Goldin-Meadow, S. (2013). Spatial language facilitates spatial cognition: Evidence from children who lack language input. *Cognition, 127*, 318–330. doi:10.1016/j.cognition.2013.01.003

Gentner, D., & Rattermann, M. J. (1991). Language and the career of similarity. In S. A. Gelman & J. P. Byrnes (Eds.), *Perspectives on language and thought: Interrelations in development* (pp. 225–277). New York, NY: Cambridge University Press.

Gentner, D., Rattermann, M. J., Markman, A. B., & Kotovsky, L. (1995). Two forces in the development of relational similarity. In T. J. Simon & G. S. Halford (Eds.), *Developing cognitive competence: New approaches to process modeling* (pp. 263–313). Hillsdale, NJ: Erlbaum.

Gentner, D., & Toupin, C. (1986). Systematicity and surface similarity in the development of analogy. *Cognitive Science, 10*, 277–300. doi:10.1207/s15516709cog1003_2

Gerken, L. (2006). Decisions, decisions: Infant language learning when multiple generalizations are possible. *Cognition, 98*, B67–B74. doi:10.1016/j.cognition.2005.03.003

Gervain, J., Berent, I., & Werker, J. F. (2012). Binding at birth: The newborn brain detects identity relations and sequential position in speech. *Journal of Cognitive Neuroscience, 24*, 564–574. doi:10.1162/jocn_a_00157

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15*, 1–38. doi:10.1016/0010-0285(83)90002-6

Giurfa, M., Zhang, S., Jennett, A., Menzel, R., & Srinivasan, M. V. (2001). The concepts of "sameness" and "difference" in an insect. *Nature, 410*, 930–933. doi:10.1038/35073582

Gomez, R. L., & Gerken, L. A. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition, 70*, 109–135.

Gomez, R. L., & Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences, 4,* 178–186. doi:10.1016/s1364-6613(00)01467-4

Halford, G. S. (1992). Analogical reasoning and conceptual complexity in cognitive development. *Human Development, 35,* 193–217. doi:10.1159/000277167

Haryu, E., Imai, M., & Okada, H. (2011). Object similarity bootstraps young children to action-based verb extensions. *Child Development, 82,* 674–686. doi:10.1111/j.1467-8624.2010.01567.x

Haun, D. B. M., & Call, J. (2009). Great apes' capacities to recognize relational similarity. *Cognition, 110,* 147–159. doi:10.1016/j.cognition.2008.10.012

Hespos, S. J., & Baillargeon, R. (2001). Infants' knowledge about occlusion and containment events: A surprising discrepancy. *Psychological Science, 12,* 141–147. doi:10.1111/1467-9280.00324

Hespos, S. J., & Baillargeon, R. (2006). Decalage in infants' knowledge about occlusion and containment events: Converging evidence from action tasks. *Cognition, 99,* B31–B41. doi:10.1016/j.cognition.2005.01.010

Hespos, S. J., Grossman, S. R., & Saylor, M. M. (2010). Infants' ability to parse continuous actions: Further evidence. *Neural Networks, 23,* 1026–1032. doi:10.1016/j.neunet.2010.07.010

Hespos, S. J., & Piccin, T. B. (2009). To generalize or not to generalize: Spatial categories are influenced by physical attributes and language. *Developmental Science, 12,* 88–95. doi:10.1111/j.1467-7687.2008.00749.x

Hespos, S. J., & Spelke, E. S. (2004). Conceptual precursors to language. *Nature, 430,* 453–456. doi:10.1038/nature02634

Hochmann, J., Mody, S., & Carey, S. (2013). *Same and different relations in match- and mismatch-to-sample.* Presented at the Cognitive Development Society, Memphis, TN.

James, W. (1890). *The principles of psychology.* Cambridge, MA: Harvard University Press.

Johnson, S. P., Fernandes, K. J., Frank, M. C., Kirkham, N., Marcus, G., Rabagliati, H., & Slemmer, J. A. (2009). Abstract rule learning for visual sequences in 8- and 11-month-olds. *Infancy, 14,* 2–18. doi:10.1080/15250000802569611

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition, 83,* B35–B42. doi:10.1016/s0010-0277(02)00004-5

Kotovsky, L., & Gentner, D. (1996). Comparison and categorization in the development of relational similarity. *Child Development, 67,* 2797–2822. doi:10.2307/1131753

Kuehne, S. E., Gentner, D., & Forbus, K. D. (2000). Modeling infant learning via symbolic structural alignment. In L. R. Gleitman & A. K. Joshi (Eds.), *Proceedings of the Twenty-Second Annual Conference of the Cognitive Science Society* (pp. 286–291). Mahwah, NJ: Erlbaum.

Maguire, M. J., Hirsh-Pasek, K., Golinkoff, R. M., & Brandone, A. C. (2008). Focusing on the relation: Fewer exemplars facilitate children's initial verb learning and extension. *Developmental Science, 11,* 628–634. doi:10.1111/j.1467-7687.2008.00707.x

Marcus, G. F. (2000). Pabiku and Ga Ti Ga: Two mechanisms infants use to learn about the world. *Current Directions in Psychological Science, 9,* 145–147. doi:10.1111/1467-8721.00080

Marcus, G. F., Fernandes, K. J., & Johnson, S. P. (2007). Infant rule learning facilitated by speech. *Psychological Science, 18,* 387–391. doi:10.1111/j.1467-9280.2007.01910.x

Marcus, G. F., Vijayan, S., Rao, S. B., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science, 283,* 77–80. doi:10.1126/science.283.5398.77

Markman, A. B., & Gentner, D. (1993). Structural alignment during similarity comparisons. *Cognitive Psychology, 25,* 431–467. doi:10.1006/cogp.1993.1011

McDonough, L., Choi, S., & Mandler, J. M. (2003). Understanding spatial relations: Flexible infants, lexical adults. *Cognitive Psychology, 46,* 229–259. doi:10.1016/s0010-0285(02)00514-5

Mix, K. S. (2008). Children's equivalence judgments: Crossmapping effects. *Cognitive Development, 23,* 191–203. doi:10.1016/j.cogdev.2007.03.001

Namy, L. L., & Gentner, D. (2002). Making a silk purse out of two sow's ears: Young children's use of comparison in category learning. *Journal of Experimental Psychology: General, 131,* 5–15. doi:10.1037/0096-3445.131.1.5

Paik, J. H., & Mix, K. S. (2006). Preschoolers' use of surface similarity in object comparisons: Taking context into account. *Journal of Experimental Child Psychology, 95,* 194–214. doi:10.1016/j.jecp.2006.06.002

Palmer, S. E. (1977). Hierarchical structure in perceptual representation. *Cognitive Psychology, 9,* 441–474. doi:10.1037/e666602011-014

Penn, D. C., Holyoak, K. J., & Povinelli, D. J. (2008). Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. *Behavioral and Brain Sciences, 31,* 109–130. doi:10.1017/s0140525x08003798

Pinto, J. P. (1996). *Xhab: Experimental control software for MS-DOS* (Version 6.5) [Computer software]. Palo Alto, CA: Author.

Premack, D. (1983). The codes of man and beasts. *Behavioral and Brain Sciences, 6,* 125–167. doi:10.1017/s0140525x00015077

Quinn, P. C., Cummins, M., Kase, J., Martin, E., & Weissman, S. (1996). Development of categorical representations for above and below spatial relations in 3- to 7-month-old infants. *Developmental Psychology, 32,* 942–950. doi:10.1037/0012-1649.32.5.942

Richland, L. E., Morrison, R. G., & Holyoak, K. J. (2006). Children's development of analogical reasoning: Insights from scene analogy problems. *Journal of Experimental Child Psychology, 94,* 249–271. doi:10.1016/j.jecp.2006.02.002

Robinson, J. S. (1955). The sameness-difference discrimination problem in chimpanzee. *Journal of Comparative* *and Physiological Psychology, 48,* 195. doi:10.1037/h0042463

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274,* 1926–1928. doi:10.1126/science.274.5294.1926

Saffran, J. R., Pollak, S. D., Seibel, R. L., & Shkolnik, A. (2007). Dog is a dog is a dog: Infant rule learning is not specific to language. *Cognition, 105,* 669–680. doi:10.1016/j.cognition.2006.11.004

Sinha, P., Balas, B., Ostrovsky, Y., & Russell, R. (2006). Face recognition by humans: Nineteen results all computer vision researchers should know about. *Proceedings of the Institute of Electrical and Electronics Engineers, 94,* 1948–1962. doi:10.1109/jproc.2006.884093

Smirnova, A., Zorina, Z., Obozova, T., & Wasserman, E. (2015). Crows spontaneously exhibit analogical reasoning. *Current Biology, 25,* 256–260.

Smith, J. D., Redford, J. S., Haas, S. M., Coutinho, M. V., & Couchman, J. J. (2008). The comparative psychology of same-different judgments by humans (*Homo sapiens*) and monkeys (*Macaca mulatta*). *Journal of Experimental Psychology: Animal Behavior Processes, 34,* 361–374. doi:10.1037/0097-7403.34.3.361

Smith, L. B. (1984). Young children's understanding of attributes and dimensions: A comparison of conceptual and linguistic measures. *Child Development, 55,* 363–380. doi:10.2307/1129949

Thibaut, J.-P., French, R., & Vezneva, M. (2010). The development of analogy making in children: Cognitive load and executive functions. *Journal of Experimental Child Psychology, 106,* 1–19. doi:10.1016/j.jecp.2010.01.001

Thompson, R. K., & Oden, D. L. (1995). A profound disparity revisited: Perception and judgment of abstract identity relations by chimpanzees, human infants, and monkeys. *Behavioral Processes, 35,* 149–161. doi:10.1016/0376-6357(95)00048-8

Thompson, R. K. R., & Oden, D. L. (2000). Categorical perception and conceptual judgments by nonhuman primates: The paleological monkey and the analogical ape. *Cognitive Science, 24,* 363–396. doi:10.1207/s15516709cog2403_2

Thompson, R. K. R., Oden, D. L., & Boysen, S. T. (1997). Language-naive chimpanzees (*Pan troglodytes*) judge relations between relations in a conceptual matching-to-sample task. *Journal of Experimental Psychology: Animal Behavior Processes, 23,* 31–43. doi:10.1037/0097-7403.23.1.31

Tyrrell, D. J., Stauffer, L. B., & Snowman, L. G. (1991). Perception of abstract identity/difference relationships by infants. *Infant Behavior and Development, 14,* 125–129. doi:10.1016/0163-6383(91)90059-2

Tyrrell, D. J., Zingardo, M. C., & Minard, K. L. (1993). Learning and transfer of identity difference relationships by infants. *Infant Behavior and Development, 16,* 43–52. doi:10.1016/0163-6383(93)80027-6

Walker, C. M., & Gopnik, A. (2014). Toddlers infer higher-order relational principles in causal learning. *Psychological Science, 25,* 161–169. doi:10.1177/0956797613502983

Wang, S.-H., & Baillargeon, R. (2008). Can infants be "taught" to attend to a new physical variable in an event category? The case of height in covering events. *Cognitive Psychology, 56,* 284–326. doi:10.1016/j.cogpsych.2007.06.003

Wasserman, E. A., & Young, M. E. (2010). Same–different discrimination: The keel and backbone of thought and reasoning. *Journal of Experimental Psychology: Animal Behavior Processes, 36,* 3–22. doi:10.1037/a0016327

Wright, A. A., & Katz, J. S. (2006). Mechanisms of same/different concept learning in primates and avians. *Behavioural Processes, 72,* 234–254. doi:10.1016/j.beproc.2006.03.009

Zentall, T. R., Wasserman, E. A., Lazareva, O. F., Thompson, R. K. R., & Rattermann, M. J. (2008). Concept learning in animals. *Comparative Cognition & Behavior Review, 3,* 13–45. doi:10.3819/ccbr.2008.30002

Child Development, September/October 2015, Volume 86, Number 5, Pages 1406–1418

Parents' Talk About Letters With Their Young Children

Rebecca Treiman, John Schmidt, and Kristina Decker
Washington University in St. Louis

Sarah Robins
University of Kansas

Susan C. Levine and Özlem E. Demir
University of Chicago

A literacy-related activity that occurs in children's homes—talk about letters in everyday conversations—was examined using data from 50 children who were visited every 4 months between 14 and 50 months. Parents talked about some letters, including those that are common in English words and the first letter of their children's names, especially often. Parents' focus on the child's initial was especially strong in families of higher socioeconomic status, and the extent to which parents talked about the child's initial during the later sessions of the study was related to the children's kindergarten reading skill. Conversations that included the child's initial were longer than those that did not, and parents presented a variety of information about this letter.

Learning to read is crucial for success in school and life. Consequently, researchers, educators, and policy makers are interested in finding out why some children learn to read more easily than others. Part of the answer may lie in the literacy-related activities that children participate in at home, before formal reading instruction begins. These activities may include being read to by their parents, learning to spell their names, and playing with magnets in the shapes of letters. Children who are reported by their parents to engage in such activities infrequently are on average less successful in learning to read than children who are reported to engage in them often (e.g., Burgess, Hecht, & Lonigan, 2002; Christian, Morrison, & Bryant, 1998). Shared book reading is the most studied aspect of the home literacy environment, but researchers have suggested that the construct of home literacy be expanded to include other activities (e.g., Phillips & Lonigan, 2009). Here, we focus on one potentially important but understudied activity—parents' talk about letters of the alphabet with their young children—and how this varies across families and relates to children's later reading performance.

When considering how children's early experiences at home set the stage for reading, it is important to ask what young readers need to learn. One important skill that must be mastered during the first few years of formal schooling is the ability to sound out individual words from text, that is, to decode. Decoding, in turn, rests on letter knowledge and phonological skills (Lonigan, Burgess, & Anthony, 2000). Among the many activities that are included in questionnaire studies of literacy-related activities in homes, those that seem to be most closely related to children's later decoding skill include parents' engagement with children in activities involving letters of the alphabet and reading and writing words (Evans, Shaw, & Bell, 2000; Hood, Conlon, & Andrews, 2008; Sénéchal & Lefevre, 2002; Skibbe, Bindman, Hindman, Aram, & Morrison, 2013; Sylva et al., 2011). The frequency of shared book reading appears to be more closely linked to vocabulary and listening comprehension than to decoding (Sénéchal & Lefevre, 2002). Thus, when looking at how experiences at home set the stage for early reading and decoding, parent talk about letters is a critical aspect to examine.

Although the studies just cited suggest that parents' talk and teaching about letters is related to

---

Kristina Decker is now at University of Memphis.

This research was supported by NIH Grants HD040605 and HD051610.

Correspondence concerning this article should be addressed to Rebecca Treiman, Washington University in St. Louis, Campus Box 1125, One Brookings Drive, St. Louis, MO 63130–4899. Electronic mail may be sent to rtreiman@wustl.edu.

© 2015 The Authors
Child Development © 2015 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2015/8605-0007
DOI: 10.1111/cdev.12385

children's decoding skills, these studies have some limitations. Many studies have asked parents whether and how often they engage in various activities. However, parents may inflate their reports of how often they perform socially valued activities to present themselves as good parents. In addition, the detail of the data that can be elicited through questionnaires is limited. Another concern is that most studies have examined the later preschool years, when children are around 4 and 5 years old. Earlier parental input may be important too, but only a few studies have examined literacy-related activities in the homes of toddlers (e.g., Burgess et al., 2002; Neumann, Hood, & Neumann, 2009).

In an attempt to overcome these limitations, one line of research (Robins, Ghosh, Rosales, & Treiman, 2014; Robins & Treiman, 2009; Robins, Treiman, & Rosales, 2014; Robins, Treiman, Rosales, & Otake, 2012) has examined how U.S. parents talk with their children about literacy-related matters by using data from the Child Language Data Exchange System (CHILDES), a computerized repository containing transcripts of communication in spoken language (MacWhinney, 2000). Robins and colleagues found that such talk occurs with children as young as 1–2 years of age. For example, a parent might mention the letters on the license plate of a toy car while playing with the child. The researchers found that parents emphasized some letters of the alphabet over others by using some letter names more frequently (Robins, Treiman, et al., 2014). Moreover, certain aspects of parents' letter talk changed across the toddler and preschool years. For example, the frequency with which parents talked about specific letters appeared to be more closely tied to the frequency of those letters in English words when children were 4 or 5 years old than when they were 1 or 2 years old (Robins, Treiman, et al., 2014). This change may reflect a greater emphasis on spelling words and associating letters with words as children get older. Other aspects of parent letter talk did not appear to change across the preschool and toddler years. Throughout this period, for example, parents often talked about A, B, and C—the first three letters of the sequence and the ones that are often used as a label for the alphabet (Robins, Treiman, et al., 2014). Although these studies provide useful information about a potentially important but understudied literacy-related activity that occurs in homes, use of data from CHILDES has its own limitations. The data collection procedures differed across the studies in CHILDES; for example, an experimenter supplied toys or books in some studies but not others. Some children were studied longitudinally and others were not, information about the literacy outcomes of the children is not available, and information about the family's socioeconomic status (SES) is available only for some families.

In the present study, we analyzed parents' talk about letters using data from a longitudinal study that collected extensive information about children and families. The families in this study, the Chicago Language Development Project, were chosen to be representative of the greater Chicago area in ethnicity and income. The families were visited in their homes approximately every 4 months starting from when the target child was 14 months old. At each visit, the caregiver was videotaped interacting with the child. We examined the amount of letter talk that parents engaged in with their children from the 14-month through the 50-month home visits and the nature of that talk, such as which letters parents most often talked about. Moreover, we asked whether the amount and nature of parent letter talk before children enter kindergarten related to the children's decoding skills at the end of kindergarten. To help determine whether any relations were specific to decoding, we also examined children's kindergarten performance on a standardized test of receptive vocabulary.

A particular focus of the present study was on parents' talk about the first letters of their children's names. Children's names, especially the first letters of the names, play an important role in early literacy development (e.g., Both-de Vries & Bus, 2008, 2010; Levin & Aram, 2005; Levin, Both-de Vries, Aram, & Bus, 2005). When asked to identify visually presented letters, for example, 4- to 6-year-olds tend to perform well on the first letter of their given name (e.g., Treiman & Broderick, 1998; Treiman, Kessler, & Pollo, 2006). Children's good performance on the first letter of their name might reflect, in part, greater exposure to this letter. However, previous studies have not provided the data needed to test this idea. For example, Robins and colleagues (Robins, Ghosh, et al., 2014; Robins, Treiman, et al., 2014) did not include the status of a letter in the child's name in their statistical models of parent letter use because the names of a number of the children in CHILDES are not available. Moreover, even though some early childhood educators have suggested that name-focused activities play an important function in teaching children about letters, reading, and writing (e.g., Kirk & Clark, 2005), no quantitative studies have examined whether parental talk about the letters in their children's names during the preschool years is related to the children's later decoding skills. We addressed these questions in Study 1, which examined parents' letter use during the 10 home visits, and Study 2, which looked in depth at letter-related conversations.

Study 1

Method

Participants

We used data from 50 children and their parents. They were drawn from a sample of families in the Chicago, Illinois area who were participating in a longitudinal study of children's language development. Families were recruited via direct mailings to approximately 5,000 families living in targeted zip codes and an advertisement in a free monthly magazine for parents. Interested parents were interviewed about their background characteristics, and 64 families who were representative of the greater Chicago area in ethnicity and income were selected. In all of the families, parents spoke English at home as the primary language. For the present study, we used data from families that remained in the study at the end of the child's kindergarten year and where data were available on a reading measure that was administered to the child at this time. Data from 4 of the 54 families that fit this description were not included because both parents shared the primary caregiving role. The language input to the child was in some sessions divided between the two parents and so was not comparable in some ways to the input from a single parent. In the 50 families that formed the final sample, the primary caregiver was the mother in 49 and the father in 1. The children included 27 boys and 23 girls, 37 of whom were reported to be White, 9 African American, and 4 of two or more races. Five of the children were reported to be Hispanic.

Information about the education level of the primary caregiver and the family's income was collected categorically in a questionnaire that was given at or before the first home visit. Each category for education was assigned a value equivalent to years of education. For example, completion of high school received a value of 12 and completion of an undergraduate degree received a value of 16. The categories for family income, which ranged from less than $15,000 to over $100,000 per year, were transformed into a scale by using the midpoints of the incomes in each category except the highest, which was coded as $100,000. Table 1 shows the mean values on these scales for the families in the study. Education and income were positively correlated ($r = .40$, $p = .004$). As in several previous studies using data from the Chicago Language Development Project (Gunderson & Levine, 2011; Levine, Suriyakham, Rowe, Huttenlocher, & Gunderson, 2010; Rowe, Raudenbush, & Goldin-Meadow, 2012), we used principal components analysis to combine education and income into a composite measure of SES with a mean of 0 and a standard deviation of 1.0. Families with higher scores on this composite measure had higher incomes and primary caregivers with higher levels of education.

Procedure

Home visits. We analyzed data from home visits that took place when each child was approximately 14, 18, 22, 26, 30, 34, 38, 42, 46, and 50 months of age. The visits, which began in 2002, were conducted by research assistants, each of whom continued with a family over a series of visits. At each visit, the research assistant videotaped the parent–child dyad for a target length of 90 min. Not all sessions exactly met this target due to variation in parents' schedules or experimenter error, but 92% of the visits were within 4 min of it and the mean length was 88.5 min. The goal was to obtain a picture of typical parent–child interactions, and so the research assistant did not bring toys but instead asked parents to interact with their child as they normally would. The activities in which parents and children engaged varied, but typical sessions included activities such as playing with toys and eating. All caregiver speech to the child and all child speech in the videotaped sessions were transcribed; singing was not transcribed. The unit of transcription was the utterance, which was defined as a sequence of words that was preceded and followed by a pause, a change in a conversational turn, or a change in intonation pattern. Transcrip-

Table 1
Information About Families in Study 1

Variable	Mean	*SD*	Range
Family income	$60,300	$31,023	Less than $15,000–over $100,000
Years of education of primary caregiver	15.7	2.1	Did not complete high school–completed advanced degree

tion reliability was established by having a second individual transcribe 20% of each transcriber's videotapes. Reliability was assessed at the utterance level and was achieved when coders agreed on 95% of transcription decisions.

We counted the number of uses of each letter in each session by each parent, whether the letter form was being pointed out visually (e.g., "All them Gs" when referring to images of the letter G in a television program), discussed as part of a spelling (e.g., "It begins with a P" in a discussion of the word *plank*), or mentioned for its sound (e.g., "/wə/,/wə/, W's for Wendy") or in some other manner. For *A* and *I*, we counted uses that were letter names and excluded those that were the article or the pronoun. Cases in which a letter name was part of a word, such as *TV* and *ABC soup*, were also excluded. For sessions that were not exactly 90 min, we adjusted the number of uses of each letter so that it reflected what it would have been had the session been 90 min, assuming a linear relation between session length and letter talk. We also tabulated the total number of word tokens that parents used in each session, adjusting it in a similar manner. Data on parent talk were not available from 7 of the potential 500 home visits for this study because the visit could not be scheduled in a timely manner or because the parent was not at home during the visit. Five families had one missing visit and one family had two. We describe how missing data were treated when presenting the individual analyses.

Kindergarten tests. During the spring of the children's kindergarten year, when the children were on average 75 months of age (*SD* = 5.3), they were given the Woodcock–Johnson Letter-Word Identification subtest, which requires them to name letters and read words aloud, and the Word Attack subtest, which requires them to pronounce nonsense words (Woodcock, McGrew, & Mather, 2001). We calculated the child's standardized score on the basic reading cluster, which is based on the scores for both the Letter-Word Identification and Word Attack subtests. The Peabody Picture Vocabulary Test (PPVT–III; Dunn & Dunn, 1997) was also given in kindergarten to all but one of the children. This task requires children to point to one of four pictures that corresponds to an orally presented word.

Results

Total Amount of Letter Talk

In our first set of analyses, we examined the amount of letter talk produced by parents, asking whether the total amount of letter talk across the 10 sessions was related to children's kindergarten reading performance. The average number of letter tokens per parent per 90-min session was 8.3 (*SD* = 9.5). Parents varied substantially in their letter use. For example, two parents produced no letter names in any session and one parent averaged 49.3 letters per session. Because parent letter use was positively skewed, subsequent analyses were performed on log transformed data (natural log of number of letter uses + 1).

To determine whether the amount of parent letter talk changed as children grew older, we carried out a repeated measures analysis of variance (ANOVA) using data from the 44 families that had data from all 10 sessions. We found a significant effect of session, $F(6.2, 265.2) = 2.47$, $p = .023$; the Greenhouse–Geisser correction was used because of a lack of sphericity. This effect occurred because parents were increasingly likely to talk about letters as children grew older. This was true even though, according to another ANOVA, the number of words that a parent spoke that were not letter names did not vary as a function of session ($p = .75$). The percentage of all parent word tokens that were letter names was 0.18 during Sessions 1–5, increasing to 0.26 during Sessions 6–10.

A regression analysis on the data from these same 44 families showed that the composite measure of SES that was described earlier contributed significantly to the prediction of children's kindergarten reading scores ($\beta = .30$, $p = .045$). When the total amount of parent letter use across the 10 sessions was added in a second step, this variable did not make a significant additional contribution.

Nature of Letter Talk

Although the results so far show that the amount of parental letter talk increased across the toddler and preschool years, they do not show whether the letters that parents talked about changed. The analyses reported in this section were designed to determine whether talk about letters with certain characteristics became increasingly common as children get older. We also asked whether parental talk about letters with certain characteristics was related to the family's SES and the child's kindergarten reading and vocabulary.

We used mixed model analyses to examine the characteristics of letters and children that were associated with letter use because such analyses are well suited to examining both types of characteristics simultaneously. For example, one letter-related factor is the frequency of the letter in English words and one child-related factor is the child's age at the home visit. Another factor of interest is a joint function of the letter and the child: whether the letter is the first letter of the child's given name. By including all factors in the same analysis, we can determine whether each factor was associated with parent letter use after the influences of other factors were statistically controlled. For example, use of the first letter of the child's name would be expected to occur at high rates in the parents of Ann and Arthur, whose names begin with a letter that is common in English words, and at low rates in the parents of Quinn and Zoe, whose names begin with uncommon letters. Using mixed model analyses, we can control for such differences. A further advantage of this statistical approach is that we use the child's actual age at each home visit rather than the target age, accounting for the fact that home visits did not always occur on the exact day that a child reached the target age. Also, rather than omitting data from the 12% of families who missed one or two sessions, as in the analyses of the amount of letter talk reported above, or rather than imputing data, we omitted from the mixed model analyses just data from sessions that were missed by a particular family (1.4% of all sessions). The analyses were conducted using R version 3.0.2 (R Core Team, 2013), using the package lme4 (Bates, Maechler, Bolker, & Walker, 2013) to perform the mixed model analyses and the package lmerTest (Kuznetsova, Brockhoff, & Christensen, 2014) to calculate *p* values based on Satterthwate's approximation. We centered continuous dependent variables prior to analysis, and the models included by-participant random intercepts.

Table 2 shows the mean and standard deviation of each fixed factor in our mixed model analyses. One factor was whether the letter was A, B, or C: three letters that parents tend to use often according to earlier research (Robins, Ghosh, et al., 2014; Robins, Treiman, et al., 2014). We also included the frequency of the letter in written materials that are designed for young children, specifically, the number of occurrences of the letter across the 6,231 words that appear in a survey of written materials in English for kindergarten and first-grade children (Zeno, Ivens, Millard, & Duvvuri, 1995). Because letter frequency as calculated in this manner was moderately skewed, we used a square root transformation in all analyses. We coded a letter as the child's initial if it was the first letter of the name that the parent most often called the child, either the full name or a nickname. We did not code letters for their occurrence later in children's first names or their last names because studies of children's ability to identify visually presented letters show small or no effects of the child's name in these cases (Treiman & Broderick, 1998). Also, some of the children in the study were called by several versions of a name, such as Jay and Jason; the versions typically differed in some of the later letters but not the first one. Other factors were the child's age at the time of a home visit and the child's kindergarten reading score. Inclusion of this latter factor, which was treated as a continuous variable, allowed us to ask whether parent talk about letters during the toddler and preschool years differed for children who became better and poorer readers in kindergarten. This analytic strategy is similar to that used in other studies aiming to identify early predictors of later life outcomes, such as studies asking whether social behavior differs in infants who are later diagnosed as autistic and those who are not (Werner, Dawson, Osterling, & Dinno, 2000). A final factor was the composite measure of family SES.

We used a step-up strategy for model building (see West, Welch, & Galecki, 2007). We first built a model that included the first three variables shown in Table 2, which are characteristics of the letter that was uttered (log-likelihood = −5,971.0, *df* = 6). In a second model, we added child age and its interaction with each letter-related variable (log-likelihood = −5,936.3, *df* = 10). This second model accounted for significantly more variance than the first model by a likelihood ratio test, $\chi^2(4) = 69.33$, $p < .001$. In a third step, we added SES and its

Table 2
Variables Included in Mixed-Model Analyses of Parent Letter Use in Study 1

Variable	*M*	*SD*	Range
ABC (whether letter is A, B, or C)	0.12	0.32	0, 1
Letter frequency in children's book corpus (square root transformed)	295.12	153.26	28.16 to 591.58
Initial (whether letter is first letter of child's given name)	0.14	0.19	0, 1
Age (years)	2.68	0.96	1.13 to 4.37
Kindergarten reading (standard score)	114.90	17.06	77 to 160
SES	0.00	1.00	−2.64 to 1.41

Note. SES refers to composite score on measure of socioeconomic status.

interactions with each term in the second model (log-likelihood = –5,907.7, $df = 18$). The amount of variance accounted for by the model again increased significantly, $\chi^2(8) = 57.17$, $p < .001$. A fourth model added the child's kindergarten reading score and its interactions with each term in the third model (log-likelihood = –5,878.9, $df = 34$). Again, there was a significant increase in the amount of variance that was explained, $\chi^2(16) = 57.72$, $p < .001$. We simplified this fourth model by excluding the interactions of ABC that involved age, kindergarten reading, and SES and the interactions of letter frequency that involved kindergarten reading and SES. These interactions were not significant, and removing them did not significantly weaken the model, $\chi^2(15) = 16.72$, $p = .34$. The results of the fixed effects are interpreted according to this more parsimonious fifth model (log-likelihood = –5,887.2, $df = 19$). Table 3 provides information about the fixed effects in this final model.

The main effect of ABC in the final model shows that parents were more likely to say A, B, and C than expected on the basis of other factors. Of the letter names that parents produced, 17% were either A, B, or C. The ABC variable did not interact significantly with any other variables in the model. That is, the priority for A, B, and C remained constant over the 14- to 50-month period, and it did not vary with the future reading ability of the child.

The final model in Table 3 also shows a main effect of letter frequency. Controlling for other factors, parents were more likely to talk about letters that are frequent in printed materials for children than about letters that are less common. The interaction between letter frequency and age was statistically significant, with the effect of letter frequency increasing as children got older. When we analyzed the results for individual sessions, using a version of the final model that omitted the effects involving child age, we found that Session 4 (26 months) was the first to show a statistically significant effect of letter frequency. The effect was consistently significant starting at Session 7 (38 months).

The final model showed a significant main effect of child's initial, such that parents were more likely to say the first letter of the child's name than expected on the basis of other factors. Of the letter names said by parents, 9% were the first letter of the child's name. The significant interaction of child's initial and age reflects the fact that during the age range covered by our study, parents were increasingly likely to say the child's initial as compared to other letters as children got older. The focus on the child's initial was particularly strong in parents of children who went on to become good readers, as shown by the significant three-way interaction involving child's initial, age, and kindergarten reading.

Table 3

Predictors in Final Mixed-Model Analysis of Parent Letter Use in Study 1

Effect	Coefficient	*SE*	*p*
ABC	5.86×10^{-2}	1.05×10^{-2}	< .001
Letter frequency	1.32×10^{-4}	2.20×10^{-5}	< .001
Child's initial	1.24×10^{-1}	1.81×10^{-2}	< .001
Age	1.89×10^{-2}	3.68×10^{-3}	< .001
Kindergarten reading	-5.87×10^{-6}	1.18×10^{-3}	.996
SES	2.36×10^{-2}	2.08×10^{-2}	.263
Letter Frequency × Age	6.51×10^{-5}	2.28×10^{-5}	.004
Child's Initial × Age	4.55×10^{-2}	1.82×10^{-2}	.012
Child's Initial × Kindergarten Reading	1.53×10^{-3}	1.09×10^{-3}	.158
Child's Initial × SES	7.31×10^{-2}	1.92×10^{-2}	< .001
Age × Kindergarten Reading	5.89×10^{-4}	2.21×10^{-4}	.008
Age × SES	2.25×10^{-2}	3.84×10^{-3}	< .001
Kindergarten Reading × SES	-3.77×10^{-4}	1.02×10^{-3}	.713
Child's Initial × Age × Kindergarten Reading	2.72×10^{-3}	1.08×10^{-3}	.012
Child's Initial × Kindergarten Reading × SES	1.99×10^{-3}	9.34×10^{-4}	.033
Age × Kindergarten Reading × SES	1.02×10^{-3}	1.88×10^{-4}	< .001

Note. SES refers to composite score on measure of socioeconomic status.

We carried out several follow-up analyses to help understand the important three-way interaction involving child's initial, age, and kindergarten reading. Analyzing the results for individual sessions, we found that Session 6 (34 months) was the first to show a significant interaction between child's initial and kindergarten reading that reflected more use of the child's initial in parents of children who went on to become good readers than in children who went on to become less good readers. When we examined the pooled data from before this point, namely, from Sessions 1 to 5 (14–30 months), using a model that omitted the variable of child age, we found a significant effect of child's initial such that parents used this letter more often than expected on the basis of other factors. However, child's initial did not interact with kindergarten reading score. Of the letter names that parents used during the first five sessions, 5% were the child's initial among parents of children who went on to score above the median on the kindergarten reading test and 8% were the child's initial among parents of children who went on to score below the median. That is, parents of children who became good readers in kindergarten did not show a stronger focus on the child's initial during Sessions 1–5 than parents of children who became poorer readers in kindergarten. In the pooled data from Sessions 6–10 (34–50 months), there was a significant interaction between child's initial and kindergarten reading such that whether a letter was the first letter of the child's name was more strongly associated with parental letter use for parents of children who scored above the median on the kindergarten reading test than parents of children who scored below the median. During this period, 13% of the letter names produced by parents of children who became good readers were the child's initial, as compared to 7% of the letter names produced by parents of children who became less good readers.

Another follow-up analysis was conducted to determine whether parental focus on the child's initial was related to kindergarten vocabulary performance as measured by the PPVT. When we replaced the kindergarten decoding variable with the standard score on the PPVT, the interaction between child's initial, age, and vocabulary was not significant ($p = .44$). Thus, although a parent's tendency to talk more about their child's initial than anticipated on the basis of other variables when the child was around 3 and 4 years old was related to the child's decoding skills in kindergarten, even after controlling for other factors, it did not appear to be related to the children's vocabulary in kindergarten.

Returning to the results reported in Table 3, there was no significant effect of SES. Over the entire 14- to 50-month period, lower SES parents were no less likely to talk about letters than higher SES parents. However, there was a significant two-way interaction between child's initial and SES. This interaction reflects the fact that, of the letters produced by parents, the proportion that were the first letter of the child's name was larger in parents who were above the median in SES, 10%, than in parents who were below the median in SES, 6%. The focus on the child's initial was especially strong in higher SES parents whose children became good readers, as shown by the three-way interaction involving child's initial, kindergarten reading, and SES.

The results in Table 3 reveal a significant main effect of age. As also shown by the ANOVAs reported in the section on the amount of parent letter talk, parents were increasingly likely to talk about letters as their children got older. The significant two-way interaction between age and kindergarten reading arose because the increase in letter talk with age was greater in parents of children who became good readers than in parents of children who became poor readers. Moreover, as shown by the two-way interaction between age and SES, the increase in letter talk as children got older was greater in higher SES parents than lower SES parents. When we examined letter talk in parents of children who were above and below the median in kindergarten reading skill and above and below the median in SES, the only group for which there was no increase in parent letter talk from the first five sessions to the last five sessions was the group of parents of children who were low in both kindergarten reading and SES. This led to a significant three-way interaction involving age, kindergarten reading, and SES.

Higher SES parents tended to talk more than lower SES parents overall, with a correlation of .48 between word tokens in a session and the composite measure of SES. However, the interaction involving age and SES and the interaction involving age, SES, and kindergarten reading appear to reflect more than just differences in amount of talk, for these interactions remained significant when we added to the final model the number of parent word tokens in the session (square root transformed in order to make its distribution more normal).

Discussion

The results of Study 1 shed light on an important but little studied literacy-related activity that occurs in homes—parent talk about letters of the alphabet—in a representative sample of U.S. children. We found, in line with previous studies (Robins et al., 2012; Robins, Treiman, et al., 2014), that U.S. parents sometimes talk with their children about letters even when their children are quite young. This conversational topic became more common over the age range that we studied, 14–50 months.

Our results further show that parents talk about some letters of the alphabet more often than others. This means that children have more opportunities to learn about some letters than others. Parents mentioned A, B, and C more often than expected on the basis of other factors, probably because these letters are at the beginning of the alphabet and because they are often used as a label for it. The emphasis on A, B, and C did not change significantly across the 14- to 50-month period. Parents also talked more with their children about letters that are common in English words than about those that are less common. This effect became stronger as children got older. This change may occur as parents increasingly associate letters with words and spell words aloud. Our findings pertaining to A, B, and C and to letter frequency agree with those obtained by Robins, Treiman, et al. (2014) using the CHILDES database. We also found that parents' emphasis on A, B, and C and on letters that are frequent in the language did not differ significantly as a function of SES. This result confirms the findings that Robins, Ghosh, et al. (2014) obtained in a comparison of higher and lower SES families from CHILDES and extends them to a broader age range.

A new finding was that parents were more likely to talk about the first letter of their child's name than expected on the basis of other factors. Moreover, within the age range that we studied, the focus on the child's initial became stronger as children grew older. Previous findings suggest that the child's first name plays a special role in children's literacy development (Both-de Vries & Bus, 2008, 2010; Levin & Aram, 2005; Levin et al., 2005). The present study is the first to show a special role for the first letter of children's names in parents' speech to young children.

Another new finding is that parental focus on the first letter of the child's name was stronger in higher SES families than in lower SES families. One potential explanation for this finding is that higher SES parents are more concerned with teaching their children to write their names than are lower SES parents. However, some studies suggest that it is lower SES parents who are more likely to endorse didactic, performance-oriented instruction of specific literacy skills with young children (Lynch, Anderson, Anderson, & Shapiro, 2006; Stipek, Milburn, Clements, & Daniels, 1992). Another possible explanation for the SES difference that we observed in parents' use of the first letter of the child's name stems from results suggesting that, compared to lower SES parents, higher SES parents more often use speech to elicit conversation with their children and less often use speech to direct their children's behavior (Hoff, Laursen, & Tardif, 2002). According to this interpretation, higher SES parents may bring up the first letter of the child's name because this letter is of special interest to the child and therefore likely to spur conversation.

Differences in the amount of input that children receive about different letters may help to explain why children perform better on some letters than on others when asked to name visually presented letters, write letters to dictation, and perform similar tasks. For example, the fact that children perform better on these tasks with the first letter of their name than with other letters (e.g., Treiman & Broderick, 1998; Treiman et al., 2006) may reflect, in part, children's greater exposure to this letter at home. Amount of exposure to a word is an important determinant of vocabulary learning in general (Schwartz & Terrell, 1983), and it is likely to be an important determinant of letter name learning as well.

The results of Study 1 shed light not only on parental letter talk and how it changes across the toddler and preschool years but also on whether the amount and nature of such talk is related to children's beginning reading performance. The total amount of parent letter talk pooled across the 10 sessions of our study did not show a significant relation to kindergarten decoding performance. However, we found several indications that one particular aspect of parent letter talk—a tendency to talk more about the first letter of the child's name than about other letters—was associated with better reading outcomes even after other factors were controlled. Most notably, children whose parents talked disproportionately about the first letter of their child's name during Sessions 6 through 10, when children were between 34 and 50 months old, tended to be better decoders as 6-year-olds than other children. This effect was specific to decoding; these children did not have significantly larger receptive vocabularies. Parental emphasis on the first letter of the child's name was especially strong in children who were good readers as 6-year-olds and who also came from higher SES families.

The results involving parents' letter talk may be compared with previously published results involving parents' number talk from the Chicago Language Development Project. Levine et al. (2010) reported a significant relation between the total amount of parent number talk between 14 and 30 months and children's number knowledge at 46 months. In the present study, the relation between total amount of parent letter talk between 14 and 50 months and children's decoding skill in kindergarten was not significant. However, some types of parent letter talk at some ages were associated with kindergarten decoding performance. Likewise, Gunderson and Levine (2011) found that some types of parent talk about numbers showed a stronger relation to children's later number knowledge than did others.

Given the relation between parents' talk about the first letter of the child's name during the later sessions of the study and kindergarten reading performance, it is important to learn more about what parents are doing when they talk about the first letter of the child's name. In Study 2, therefore, we looked in detail at how and when parents talked about the first letter of their child's name. We examined the contexts in which conversations involving the child's initial occurred, and we also examined whether the conversations were initiated by parents or by children. If conversations involving the first letter of the child's name serve a teaching function, as we hypothesized, they might be especially likely to be initiated by parents. We expected that children might find such conversations particularly engaging, given how important their own names are to them. Thus, we hypothesized that conversations in which parents mentioned the child's initial would last longer than conversations in which parents mentioned only other letters. Study 2 also examined the information that parents conveyed when they used the first letter of the child's name. Did these utterances indicate that the letter in question was in the child's name (e.g., "Your name starts with J"), what we call simple name matching? Or did the utterances convey additional information about the letter (e.g., "Jar and John both start with J"), what we call complex letter activity? We hypothesized that complex letter activity would be common with older preschoolers. Because of the detailed coding that was required for Study 2, it was limited to the 30-, 42-, and 50-month visits.

Study 2

Method

Participants

The analyses of letter-related conversations involved 49 of the 50 parent–child dyads included in Study 1; 1 dyad had no letter-related conversations during the visits that were selected for inclusion in Study 2. The analyses of the nature of parents' utterances that involved the first letter of the child's name were based on the results of those 28 parents who produced such utterances during the selected visits.

Procedure

We examined all conversations in which letters were mentioned during the 30-, 42-, and 50-month sessions. We determined from the video and audio records whether each conversation occurred while the parent and child were looking at or reading a book or other written material, looking at or playing with toys, producing a written or drawn product (on paper or in some other manner), or engaging in some other activity, such as eating dinner. We refer to these conversational contexts as text, toy play, production, and other, respectively. We did not separate production into writing and drawing because the two activities were sometimes intermixed. We considered that a new conversation began when the context changed, as when a parent and child transitioned from reading a book to eating dinner, or when the focus of the conversation changed from, for example, playing with a toy to discussing a program that was being shown on television. We divided the conversations in which the parent mentioned a letter into those in which the parent said the first letter of the child's name at least once, either directly (e.g., "J is for John") or indirectly (e.g., "That's the letter that your name starts with"), and those in which the parent did not say the first letter of the child's name. We measured the duration of each letter-related conversation in number of conversational turns, that is, the number of times that the speaker changed; number of utterances by parents and children combined; and length in minutes. For each conversation, we also coded whether it was initiated by the parent or the child. The few cases in which both began speaking simultaneously were excluded from the analyses involving initiation. Reliability of the coding was assessed by having a second individual code the data from three children at each age level. The coders agreed 96% of the time on the coding of whether an utterance involved talk about letters, 86% of the time on the coding of the utterance context, and 96% of the time on the coding of conversation start and stop points.

We divided the utterances in which parents used the first letter of the child's name into three categories according to the information that parents conveyed about the letter. Simple name matching utterances were those such as "J is for John" or "J is the first letter of your name" that conveyed only that the letter in question was the first letter of the child's name and that did not provide other information about the letter or the name. Complex letter activity utterances were those in which the parent spelled the child's name or another word (e.g., "That's J O H N"; "J A R"), matched the letter to a word other than or in addition to the child's name ("J is for jar"), or discussed the sound of a letter ("It says 'juh'"). A third category consisted of all other utterances, such as "Where's the J?" or "J."

Results

The majority of letter-related conversations in which the parent mentioned the child's initial, 81%, were initiated by the parent. The percentage of letter-related conversations that were initiated by the parent for conversations in which the parent did not mention the child's initial was lower, 66%. A mixed model analysis of parent initiation that included the variables of child's initial use (parent used child's initial in the conversation vs. parent did not use child's initial), child age, and SES found a significant effect of child's initial use ($p = .008$) and no other significant effects.

Table 4 provides information about the contexts of conversations in which parents directly or indirectly mentioned the child's initial and the contexts of conversations where letters were mentioned, but not the child's initial. The interaction between conversational context and whether the letter was the child's initial was significant by a chi-square test, $\chi^2(3) = 53.34$, $p < .001$. As the results in Table 4 show, conversations in which a parent uttered the child's initial were more likely to occur in production contexts (46%) than conversations in which a parent uttered other letters (26%). A little less than half of the utterances with the child's initial in production contexts (43%, or 20% of all initial utterances) involved the parent encouraging the child to write the child's first name. The other utterances of the child's initial in production context involved other activities, such as the production of individual letters or other words. In the 42-month session, for example, one parent asked a child to write several letters. One of the requested letters was the child's initial—"Go make your best M to show Erica"—perhaps because the parent thought that the child had a good chance of success with this letter.

Table 5 provides information about the duration of conversations in which parents mentioned the child's initial and conversations in which parents mentioned only other letters. The results suggest that the conversations in which parents mentioned the child's initial were longer than those in which they did not. This impression was confirmed by mixed model analyses that included the fixed factors of child's initial use, child age, SES, and conversational context. The duration measures were log transformed to make the distributions more normal. For all three measures of conversation length, we found a significant effect of child's initial use after controlling for the other factors ($p < .001$).

Table 6 shows, for each session, the number of parent utterances that directly or indirectly used the first letter of the child's name that fell into the simple name matching, complex letter activity, and other categories. The association between category and age group was statistically significant, $\chi^2(4) = 19.27$, $p < .001$. The proportion of parent utterances that included the child's initial that fell into the simple name matching category was larger at 30 months than at 42 and 50 months. In contrast, the proportion of utterances in the complex letter activity category was larger for older children than for younger ones.

Discussion

The results of Study 1 suggest that parents' focus on the first letter of a child's name when the child is around 3 and 4 years of age is positively related

Table 4
Number of Parent Utterances Including Direct or Indirect Use of Child's Initial and Other Letters as a Function of Conversational Context, Pooling Across 30-, 42-, and 50-Month Sessions

Conversational context	Child's initial	Other letter
Text	30	549
Toy play	100	988
Production	118	566
Other	9	99

Table 5
Mean (and standard deviation) Across Participants for Various Measures of Duration for Conversations in Which Parent Explicitly or Implicitly Mentioned Child's Initial and Conversations in Which Parent Did Not Mention Child's Initial, Pooling Across 30-, 42-, and 50-Month Sessions

Measure	Child's initial included	Child's initial not included
Number of turns	18.25 (42.19)	5.40 (5.16)
Number of utterances	34.49 (64.86)	11.56 (9.30)
Length in minutes	1.11 (2.03)	0.42 (0.42)

Table 6
Number of Parent Utterances That Included Direct or Indirect Use of Child's Initial in Various Categories

Category	30 months	42 months	50 months
Simple name matching	11	4	9
Complex letter activity	12	36	54
Other	37	48	46

to the child's reading skills at the end of kindergarten, even after controlling for other factors. In Study 2, we looked in depth at parent talk about the first letter of the child's name, asking whether such talk has special properties that may make it particularly fruitful for literacy learning.

We found that conversations in which parents talked about the first letter of the child's name were typically initiated by parents rather than children. Almost half of these conversations occurred when parents and children were engaged in the production of writing: either the production of the name itself or the production of individual letters or other words. Conversations in which parents touched on the first letter of the child's name on average lasted longer than those in which parents talked only about other letters, with more participation on the part of the child. These differences may reflect children's motivation to talk about the first letter of their name, a word that is important and interesting to them (Nuttin, 1985). When parents used the first letter of the child's name, they provided information not only that this letter was at the beginning of the child's name but also, particularly with children of 42 and 50 months, information about other matters. These included letters in the child's name beyond the first letter, the sound of the first letter of the name, and other words that contained this letter. Simple name matching was less common with older children than with younger ones, as Robins, Treiman, et al. (2014) also observed.

Overall, the results of Study 2 suggest that parents who often talked about the first letter of the child's name presented a range of information about letters and words in a way that was particularly engaging for the child. The present study is correlational and descriptive rather than experimental, and we cannot draw conclusions about causation. However, it is possible that the special characteristics of parent talk about the first letter of the child's name may help to explain the association that we found in Study 1 between such talk and literacy outcomes.

General Discussion

Long before children are formally taught to read and write, their parents sometimes talk with them about letters of the alphabet while engaging in everyday activities. The results of previous studies suggest that parents' engagement with their children in activities involving letters of the alphabet and in reading and writing words correlates with children's later reading skills. With a few exceptions, however (Robins, Ghosh, et al., 2014; Robins & Treiman, 2009; Robins, Treiman, et al., 2014), previous studies relied on questionnaires and focused on older preschoolers. Here we directly observed and coded parental speech using data from a recent longitudinal study of a representative sample of children in the Chicago area. We examined the amount and nature of parents' talk about letters, asking how it changes across the toddler and preschool years, how it relates to family SES, and how it relates to children's later decoding skills.

Our results show that parents sometimes mention letters in conversations with their children even before the children are 2 years old. This aspect of the home literacy environment can be missed in questionnaires that focus on explicit teaching. For example, a mother might consider that she is playing with her child rather than teaching him when she tickles her child while using her finger to form letters on the child's back. She might therefore respond, as a number of adults do, that teaching about letters does not occur with children of this age (Burgess, 2006).

Our results show that the amount of letter talk that parents engage in with their children increases across the toddler and preschool years and that talk about certain kinds of letters changes at different rates than others. Talk about letters that are common in English words showed a substantial increase as children grow older, as also reported by Robins, Treiman, et al. (2014). A new finding was that parents were increasingly likely to talk about the first letter of the child's name over the 14- to 50-month period.

We observed differences in parent letter talk not only as a function of child age but also as a function of family SES. Notably, higher SES parents were more likely than lower SES parents to focus their letter talk on the first letter of the child's name. At first glance, this finding appears to disagree with the report of Robins, Ghosh, et al. (2014) that lower SES parents are more likely than higher SES parents to associate letters with children's names, as in "J is for Jason." However, we found in

Study 2 that parents' talk about the first letter of the child's name was by no means limited to simple name matching of this sort. This was especially true for parents of older children and for higher SES parents.

The total amount of parent letter talk over the 14- to 50-month period did not bear a statistically significant relation to children's kindergarten reading performance. However, a focus on the first letter of the child's name during the latter part of this period did. Specifically, children of parents who talked more about the first letter of the child's name when children were around 3 and 4 years of age (Sessions 5–10 of the study) tended to be better readers at the end of kindergarten than other children. This result was found after controlling for differences in the amount of talk about the first letter of the first name that would be expected on the basis of other characteristics of the letter, such as its frequency in English words.

Why did parents' talk about the first letter of their children's names when children were 3 and 4 years of age relate to the children's decoding performance in kindergarten? The results of Study 2 show that only about 20% of the utterances of the first letter of the child's first names occurred when parents were encouraging children to write their names. Thus, it does not appear to be just the act of learning to write the name that accounts for the link with later literacy development. One contributor may be young children's interest in talking about and learning about the first letter of their own names. The results of Study 2 suggest that when parents initiate conversations that touch on the first letter of their children's names, the children are particularly motivated to continue the conversations. Such conversations last longer than other letter-related conversations, extending children's opportunities to learn. Also, parents who talk about the first letter of the child's name provide information not only that this letter is at the beginning of the name but also—especially with 3- and 4-year-olds—other information about the letter, the child's name, and words other than the name. Thus, the conversations give children a variety of learning opportunities.

Our results show the value of going beyond questionnaire data in studying the early home environment and how it sets the stage for later academic performance. Naturalistic studies of parents and their interactions with their young children that include detailed coding and longitudinal follow-up measures can provide rich and valuable information about what parents of successful learners do. The results of such studies can suggest avenues for intervention that might boost the learning of all children.

References

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2013). *lme4: Linear mixed-effects models using Eigen and S4*. Retrieved from http://cran.r-project.org/package=lme4

Both-de Vries, A. C., & Bus, A. G. (2008). Name writing: A first step to phonetic writing? *Literacy Teaching and Learning*, *12*, 37–55.

Both-de Vries, A. C., & Bus, A. G. (2010). The proper name as starting point for basic reading skills. *Reading and Writing*, *23*, 173–187. doi:10.1007/s11145-008-9158-2

Burgess, S. R. (2006). The development of phonological sensitivity. In D. K. Dickinson & S. B. Neumann (Eds.), *Handbook of early literacy research* (Vol. 2, pp. 90–100). New York, NY: Guilford.

Burgess, S. R., Hecht, S. A., & Lonigan, C. J. (2002). Relations of the home literacy environment (HLE) to the development of reading-related abilities: A one-year longitudinal study. *Reading Research Quarterly*, *37*, 408–426. doi:10.1598/RRQ.37.4.4

Christian, K., Morrison, F. J., & Bryant, F. B. (1998). Predicting kindergarten academic skills: Interactions among child care, maternal education, and family literacy environments. *Early Childhood Research Quarterly*, *13*, 501–521. doi:10.1016/S0885-2006(99)80054-4

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test* (3rd ed.). Circle Pines, MN: American Guidance Service.

Evans, M. A., Shaw, D., & Bell, M. (2000). Home literacy activities and their influence on early literacy skills. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *54*, 65–75. doi:10.1037/h0087330

Gunderson, E. A., & Levine, S. C. (2011). Some types of parent number talk count more than others: Relations between parents' input and children's cardinal-number knowledge. *Developmental Science*, *14*, 1021–1032. doi:10.1111/j.1467-7687.2011.01050.x

Hoff, E., Laursen, B., & Tardif, T. (2002). Socioeconomic status and parenting. In M. H. Bornstein (Ed.), *Handbook of parenting: Biology and ecology of parenting* (Vol. 2, pp. 231–252). Mahwah, NJ: Erlbaum.

Hood, M., Conlon, E., & Andrews, G. (2008). Preschool home literacy practices and children's literacy development: A longitudinal analysis. *Journal of Educational Psychology*, *100*, 252–271. doi:10.1037/0022-0663.100.2.252

Kirk, E. W., & Clark, P. (2005). Beginning with names: Using children's names to facilitate early literacy learning. *Childhood Education*, *81*, 139–144. doi:10.1080/00094056.2005.10522257

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2014). *lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package)*. Retrieved from http://cran.r-project.org/package=lmerTest

Levin, I., & Aram, D. (2005). Children's names contribute to early literacy: A linguistic and a social perspective. In D. Ravid & H. B.-Z. Shyldkrot (Eds.), *Perspectives on language and language development* (pp. 219–239). New York, NY: Springer.

Levin, I., Both-de Vries, A., Aram, D., & Bus, A. (2005). Writing starts with own name writing: From scribbling to conventional spelling in Israeli and Dutch children. *Appied Psycholinguistics, 26,* 463–477. doi:10.1017/S0142716405050253

Levine, S. C., Suriyakham, L. W., Rowe, M. L., Huttenlocher, J., & Gunderson, E. A. (2010). What counts in the development of young children's number knowledge? *Developmental Psychology, 46,* 1309–1319. doi:10.1037/a0022101

Lonigan, C. J., Burgess, S. R., & Anthony, J. L. (2000). Development of emergent literacy and early reading skills in preschool children: Evidence from a latent-variable longitudinal study. *Developmental Psychology, 36,* 596–613. doi:10.1037//OOI2-1649.36.5.596

Lynch, J., Anderson, J., Anderson, A., & Shapiro, J. (2006). Parents' beliefs about young children's literacy development and parents' literacy behaviors. *Reading Psychology, 27,* 1–20. doi:10.1080/02702710500468708

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk* (3rd ed.). Mahwah, NJ: Erlbaum.

Neumann, M. M., Hood, M., & Neumann, D. L. (2009). The scaffolding of emergent literacy skills in the home environment: A case study. *Early Childhood Education Journal, 36,* 313–319. doi:10.1007/s10643-008-0291-y

Nuttin, J. M. (1985). Narcissism beyond Gestalt and awareness: The name letter effect. *European Journal of Social Psychology, 15,* 353–361. doi:10.1002/ejsp.2420150309

Phillips, B. M., & Lonigan, C. J. (2009). Variations in the home literacy environment of preschool children: A cluster analytic approach. *Scientific Studies of Reading, 13,* 146–174. doi:10.1080/10888430902769533

R Core Team. (2013). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.r-project.org/

Robins, S., Ghosh, D., Rosales, N., & Treiman, R. (2014). Letter knowledge in parent–child conversations: Differences between families differing in socio-economic status. *Frontiers in Psychology, 5,* 632. doi:10.3389/fpsyg.2014.00632

Robins, S., & Treiman, R. (2009). Talking about writing: What we can learn from conversations between parents and their young children. *Applied Psycholinguistics, 30,* 463–484. doi:10.1017/S0142716409090237

Robins, S., Treiman, R., & Rosales, N. (2014). Letter knowledge in parent–child conversations. *Reading and Writing, 27,* 407–429. doi:10.1007/s11145-013-9450-7

Robins, S., Treiman, R., Rosales, N., & Otake, S. (2012). Parent–child conversations about letters and pictures. *Reading and Writing, 25,* 2039–2059. doi:10.1007/s11145-011-9344-5

Rowe, M. L., Raudenbush, S. W., & Goldin-Meadow, S. (2012). The pace of vocabulary growth helps predict later vocabulary skill. *Child Development, 83,* 508–525. doi:10.1111/j.1467-8624.2011.01710.x

Schwartz, R. G., & Terrell, B. Y. (1983). The role of input frequency in lexical acquisition. *Journal of Child Language, 10,* 57–64. doi:10.1017/S0305000900005134

Sénéchal, M., & Lefevre, J.-A. (2002). Parental involvement in the development of children's reading skill: A five-year longitudinal study. *Child Development, 73,* 445–460. doi:10.1111/1467-8624.00417

Skibbe, L. E., Bindman, S. W., Hindman, A. H., Aram, D., & Morrison, F. J. (2013). Longitudinal relations between parental writing support and preschoolers' language and literacy skills. *Reading Research Quarterly, 48,* 387–401. doi:10.1002/rrq.55

Stipek, D., Milburn, S., Clements, D., & Daniels, D. H. (1992). Parents' beliefs about appropriate education for young children. *Journal of Applied Developmental Psychology, 13,* 293–310. doi:10.1016/0193-3973(92)90034-F

Sylva, K., Chan, L. L. S., Melhuish, E., Sammons, P., Siraj-Blatchford, I., & Taggart, B. (2011). Emergent literacy environments: Home and preschool influences on children's literacy development. In S. B. Neuman & D. K. Dickinson (Eds.), *Handbook of early literacy research* (Vol. 3, pp. 97–117). New York, NY: Guilford.

Treiman, R., & Broderick, V. (1998). What's in a name: Children's knowledge about the letters in their own names. *Journal of Experimental Child Psychology, 70,* 97–116. doi:10.1006/jecp.1998.2448

Treiman, R., Kessler, B., & Pollo, T. C. (2006). Learning about the letter name subset of the vocabulary: Evidence from US and Brazilian preschoolers. *Applied Psycholinguistics, 27,* 211–227. doi:10.1017/S0142716406060255

Werner, E., Dawson, G., Osterling, J., & Dinno, N. (2000). Brief report: Recognition of autism spectrum disorder before one year of age: A retrospective study based on home videotapes. *Journal of Autism and Developmental Disorders, 30,* 157–162. doi:10.1023/A:1005463707029

West, B. T., Welch, K. B., & Galecki, A. T. (2007). *Linear mixed models: A practical guide using statistical software.* Boca Raton, FL: Chapman and Hall/CRC.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock–Johnson Tests of Achievement.* Itasca, IL: Riverside.

Zeno, S. M., Ivens, S. H., Millard, R. T., & Duvvuri, R. (1995). *The educator's word frequency guide.* Brewster, NY: Touchstone Applied Science.

Child Development, September/October 2015, Volume 86, Number 5, Pages 1419–1433

Cross-Linguistic Lexical, Grammatical, and Discourse Performance on Oral Narrative Retells Among Young Spanish Speakers

Audrey Lucero
University of Oregon

This study investigated relations among microlevel and macrolevel domains of oral narrative retells within and across the languages of Spanish-speaking bilingual children. Fifty-six first and second graders ($M_{\text{age}} = 7$ years, 3 months) were assessed on a retell task in Spanish and English. Two statistical analyses were conducted: (a) correlations within and across microlevel lexical (number of different words) and grammatical (mean length of utterance-word and subordination index) domains and (b) hierarchical regression to determine the influence of microlevel domains on a macrolevel discourse score. Results highlighted a number of significant within- and cross-language correlations, and identified vocabulary as a significant predictor of macrolevel discourse scores within both languages, while grammar was a predictor within English only.

Reading comprehension is an area in which many second language learners struggle (Carlisle, Beeman, Davis, & Spharim, 1999; Muñoz, Gillam, Peña, & Gulley-Faehnle, 2003; Roth, Speece, Cooper, & De La Paz, 1996), and researchers have begun to advocate the assessment of comprehension for such students through listening and retelling (Fiestas & Peña, 2004; Miller et al., 2006; Muñoz et al., 2003). As such, oral narrative retells have been used extensively with this population in recent years (Bedore, Peña, Gillam, & Tsung-Han, 2010; Fiestas & Peña, 2004; Muñoz et al., 2003; Rojas & Iglesias, 2013). The Oral Narrative Retell Assessment (ONR) requires children to understand a story that is heard and organize a coherent, detailed retell (Bedore et al., 2010). The processing demands of oral narrative retells are therefore similar to those needed to comprehend written text (Roth et al., 1996). Because of the joint comprehension and production demands, performance on this task is thought to align closely with written literacy skills (Allen, Ukrainetz, & Carswell, 2012). Like reading, oral narrative retelling is composed of microlevel (word or sentence) and macrolevel (discourse) components. Microlevel components of retells fall within lexical and grammatical domains and are commonly measured by counting the total number of words and number of different words (NDW) used. In the grammatical domain, mean length of utterance and degree of subordination are typically measured (Allen et al., 2012; Heilmann, Miller, Nockerts, & Dunaway, 2010; Miller et al., 2006). The macrolevel domain provides a holistic picture of a child's understanding of higher order narrative skills and is assessed by measuring their ability to organize a coherent retell that includes key narrative elements. Discourse, as used in the current study, refers to "units of spoken text beyond the sentence level, and includes the ability to construct an original story and to retell a previously heard story" (Roth et al., 1996, p. 258).

Retells are considered more culturally relevant and ecologically valid than standardized tests for second language learners because they are a major component of the literacy curriculum and are not normed on monolinguals (Bedore et al., 2010; Gutiérrez-Clellen, 1998; Reese, Sparks, & Suggate, 2012). From an early age, children are expected to understand and use narrative form effectively, so assessment of such skill is appropriate and informative (Heilmann, Miller, Nockerts, & Dunaway, 2010). In addition, oral narrative tasks are not dependent on graphemic systems, so they can readily be compared across bilingual children's two languages, providing a more comprehensive picture of their language and literacy proficiency (Snow, Tabors, Nicholson, & Kurland, 1995).

One large-scale study to date has shown a robust relation between performance in oral narrative retell domains and reading proficiency in Spanish-speaking

---

Correspondence concerning this article should be addressed to Audrey Lucero, College of Education, University of Oregon, 5277 University of Oregon, Eugene, OR 97403. Electronic mail may be sent to alucero@uoregon.edu.

© 2015 The Author
Child Development © 2015 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2015/8605-0008
DOI: 10.1111/cdev.12387

second language learners (Miller et al., 2006), but little is known about interrelations among lexical, grammatical, and discourse domains of oral narrative retells within this population, especially cross-linguistically. Scholars have noted the importance of attending to oral language and literacy skills in both languages (Gottardo & Mueller, 2009; Manis, Lindsey, & Bailey, 2004; Miller et al., 2006), and doing so has both practical and theoretical benefits. Practically, it highlights the contributions of various first (L1) and second (L2) language skills to the development of literacy in both languages (Droop & Verhoeven, 2003); theoretically, it identifies which linguistic skills are universal and which may be language specific (Goodrich, Lonigan, & Farver, 2013).

Research has shown different and sometimes contradictory patterns of correlation within and across the domains of oral narrative retells (Bedore et al., 2010; Heilmann, Miller, Nockerts, & Dunaway, 2010; Simon-Cereijido & Gutiérrez-Clellen, 2009). This article builds on the scant research base by reporting findings from a study that investigated relations among microlevel and macrolevel domains within and across languages for first- and second-grade Spanish-speaking second language learners. By looking at performance across languages, I have begun to identify the patterns of skill development that may affect the acquisition of literacy in children learning a second language (Simon-Cereijido & Gutiérrez-Clellen, 2009). Attention to microlevel and macrolevel domains makes it possible to examine children's proficiency at various levels as well as to investigate relations among the various oral language skills that have been found to be significant for the development of literacy.

Oral Language Development and Assessment for Second Language Learners

The important role that oral language plays in the development of reading skill is well documented for both monolingual and bilingual children (Roth, Speece, & Cooper, 2002; Snow et al., 1995; Speece, Roth, Cooper, & De La Paz, 1999). For example, there is widespread agreement that phonological awareness is a key oral language predictor of future reading success (Bialystok, 2001; Bialystok & Herman, 1999; Carlisle et al., 1999; Durgunoglu, Nagy, & Hancin-Bhatt, 1993), and robust correlations have been reported in this domain between bilingual children's two languages (Gottardo & Mueller, 2009; Hammer & Miccio, 2006). However, phonological awareness is only one of many oral language skills, and currently less is known about within- and cross-linguistic relations among the lexical, grammatical, and discourse domains of oral language for second language learners.

Microlevel Domains

Vocabulary is the primary component of the lexical domain, and it is an area of concern for many second language learners (August, Carlo, Dressler, & Snow, 2005; Carlo et al., 2004). There is widespread agreement in the literature about the predictive validity of L1 and L2 vocabulary in reading comprehension in the same language for children as young as first grade and as old as fourth grade (Carlisle et al., 1999; Droop & Verhoeven, 2003; Lindsey, Manis, & Bailey, 2003; Proctor, August, Carlo, & Snow, 2006; Proctor, August, Snow, & Barr, 2010; Proctor, Carlo, August, & Snow, 2005). In a large-scale study of 1,531 kindergarten through third-grade Spanish-English bilingual children in transitional bilingual classrooms, Miller et al. (2006) investigated relations between oral narrative retell performance, reading comprehension, and decoding in Spanish and English. They found that the NDW used in retells—a component within the lexical domain—accounted for a significant amount of variance in both comprehension and decoding within and across languages.

A child's performance in the grammatical domain—often operationalized as awareness of morphology and syntax—is also a potential predictor of reading proficiency in second language learners of various ages (Droop & Verhoeven, 2003; Gottardo & Mueller, 2009; Lindsey et al., 2003; Manis et al., 2004), albeit a less robust one than phonological awareness or vocabulary. In the aforementioned Miller et al. (2006) study, grammatical complexity was measured using mean length of utterance, and again it was found to contribute significantly to variance in comprehension and decoding.

Discourse Domain

Macrolevel analysis of oral narrative retells focuses on discourse rather than word or sentence-level components. It is relevant to the development of literacy because the skills needed to be proficient in narrative discourse are also those needed to comprehend written text (Roth et al., 1996). Oral narrative discourse and written text share a prevalence of academic or low-frequency vocabulary, a complex syntactic style, and a coherent overall organization.

Story grammar has historically been used to measure children's discourse skills (Terry, Mills, Bingham, Mansour, & Marencin, 2013). Story grammar requires children to identify a story's setting and features of its episodic structure, including the central problem, attempts to solve the problem, and consequences or outcomes of such attempts. However, Heilmann, Miller, and Nockerts (2010) contended that measures from the story grammar tradition focus too much on the recall of specific episodes and not enough on higher level narrative skills. Therefore, they argued that such tools might not be appropriate for assessing children beyond the emergent stage of literacy development. To that end, they developed the Narrative Scoring Scheme (NSS), which incorporates features of story grammar with elements such as character development, cohesion, and the use of metalinguistic and metacognitive verbs. The NSS is scored using a combination of discrete coding criteria and researcher judgment (Heilmann, Miller, Nockerts, & Dunaway, 2010). Miller et al. (2006) used the NSS to assess the retells of children in kindergarten through third grade and found scores on it to be significantly predictive of comprehension and decoding.

Relations Among Domains

A small amount of recent research using oral narrative retells has investigated relations among microlevel lexical and grammatical components both within and across the languages of second language learners. Simon-Cereijido and Gutiérrez-Clellen (2009) conducted a study with 44 preschool- and kindergarten-aged Spanish-English bilinguals using the Mercer Mayer *Frog* books (Mayer, 1969, 1973, 1974, 1975). They found strong within-language correlations between the two domains (Spanish lexical and Spanish grammatical, $r = .71$, $p < .001$; English lexical and English grammatical, $r = .64$, $p < .001$) but no significant correlations between domains across languages (Spanish lexical and English grammatical, $r = .18$, *ns*; Spanish grammatical and English lexical, $r = .15$, *ns*). This largely corroborates findings from nonretell studies and provides support for the interdependence of domains within a language, but limited cross-linguistic relations among domains. However, Simon-Cereijido and Gutiérrez-Clellen did not calculate cross-linguistic correlations within the same domain, which could have provided information about those associations as well.

In a related study, Bedore et al. (2010) analyzed the spontaneous oral narratives of 170 kindergarten Spanish-English bilinguals using the *Frog* books (Mayer, 1969, 1974) in the interest of identifying how microlevel components that map onto lexical diversity and grammatical complexity converged with language ability. Finding relations among the domains was not the goal of the study, but correlations were nonetheless reported. Like Simon-Cereijido and Gutiérrez-Clellen (2009), Bedore et al. found highly significant correlations between lexical and grammatical domains within both languages (Spanish lexical and Spanish grammatical, $r = .65$, $p < .001$; English lexical and English grammatical, $r = .57$, $p < .001$), but no significant correlations across languages between domains (Spanish lexical and English grammatical, $r = .13$, *ns*; Spanish grammatical and English lexical, $r = .12$, *ns*). They also analyzed components across languages within domains and found evidence of cross-linguistic correlations in grammatical complexity (Spanish grammatical and English grammatical, $r = .26$, $p < .001$) but not in lexical diversity (Spanish lexical and English lexical, $r = .12$, *ns*).

Few studies to date have investigated the relations between oral narrative retell microlevel domains and discourse domain skills, and they have been conducted with only monolingual English speakers. Heilmann, Miller, and Nockerts (2010) examined the relations between lexical performance (measured using NDW), grammatical performance (based on mean length of utterance-morpheme and the subordination index [SI]), and scores on the NSS in a study of one hundred and twenty-nine 5- to 7-year-olds using the retell methodology. They sought to provide, "evidence for the role of vocabulary and grammar in the development of young school-age children's narrative organization skills" (p. 157). Findings showed correlations between lexical and grammatical domains and macrolevel narrative skills (lexical and discourse, $r = .58$, $p < .001$; grammatical and discourse, $r = .44$, $p < .001$).

Heilmann, Miller, and Nockerts (2010) also performed hierarchical regression analysis to identify the microlevel components that were the best predictors of macrolevel discourse skill. They found that children's vocabulary was the only unique predictor of discourse skill. Grammatical complexity did not add significantly to the regression equation. They therefore proposed "a special and important relation between narrative organization and vocabulary skills that emerges prior to children becoming literate" (p. 161). This finding lines up with studies that have used standardized measures of vocabulary as predictors of reading proficiency (Droop & Verhoeven, 2003; Lindsey et al., 2003; Verhoeven, 2000), thus underscoring the potential value of the oral narrative retell as a complementary assessment tool to identify the strengths and challenges that second language learners face with regards to reading.

One other study that investigated associations between microlevel and macrolevel domains of retells was conducted with preschool-aged African American children (N = 146; Terry et al., 2013). In line with Heilmann, Miller, and Nockerts's (2010) findings, they reported highly significant correlations between lexical and grammatical components and discourse skill (lexical and discourse, $r = .51$, $p < .001$; grammatical and discourse, $r = .52$, $p < .001$).

The Current Study

In this article, I report findings from a study that investigated within- and cross-linguistic relations in the lexical, grammatical, and discourse oral narrative retell performance of second language learners. Fifty-six first- and second-grade Spanish-speaking children were assessed using an oral narrative retell task in both Spanish and English. Transcripts were analyzed and compared within and across languages.

The specific research questions addressed in this study were:

1. Which microlevel components of oral narrative retells were significantly correlated within and across languages for Spanish-speaking first and second graders?
2. How well did lexical and grammatical microlevel components predict narrative discourse scores within and across languages?

Although previous findings related to the first research question have been mixed, within-language correlations among domains are generally reported to be significant (Bedore et al., 2010; Simon-Cereijido & Gutiérrez-Clellen, 2009). Therefore, it was hypothesized in the current study that there would be significant within-language correlations among components in microlevel lexical and grammatical domains. Cross-linguistic microlevel correlations were not expected.

With regard to the second research question, only a small amount of research has been conducted with monolingual children (Heilmann, Miller, Nockerts, & Dunaway, 2010; Terry et al., 2013). Nonetheless, findings have shown that vocabulary and grammar potentially predict discourse skill. Given the key role that oral language plays in the development of narrative proficiency, it was hypothesized that lexical and grammatical components would be significant predictors of NSS scores within both languages but not necessarily across languages.

Method

Participants

Data for this study were collected in four schools in two districts; both districts were located in the same mid-sized urban area in the Pacific Northwest near a large public university. District A served a diverse population of about 11,000 students, approximately 68% of whom were White and 19% Latino, as reported by the State Department of Education (Oregon, 2013). At the time of the study, 63% of students were eligible for free and reduced lunch and approximately 6% received English as a second language (ESL) services. A majority of ESL students spoke Spanish as a home language and were concentrated in certain areas of the city. Students from three schools with high concentrations of Spanish-speaking Latino students (averaging 29%) and rates of eligibility for free and reduced lunch (averaging 85%) participated. All elementary schools in the district provided instruction in English only at the time of the study.

District B enrolled approximately 16,000 students in Grades K–12, 68% of whom were White and 14% Latino. According to the State Department of Education, 43% of students were eligible for free or reduced lunch and approximately 1% received ESL services at the time of the study. As in District A, a large majority of ESL children spoke Spanish as a home language and attended a handful of elementary schools. The site for this study was a dual language immersion school with a considerably higher percentage of Latino students (41%) and those eligible for free and reduced lunch (68%) than the district average. In this school, literacy instruction in Grades K–2 was conducted in the home language of the child (as determined by school criteria and initial language assessment), so all participating children received literacy instruction in Spanish at the time of the study. The inclusion of children in different language of instruction (LOI) contexts was intended to be representative of the primary grade Spanish-speaking population of the United States in that some children possessed strong native oral language and reading skills, and others did not (Proctor et al., 2005).

Fifty-six children participated in this study, all of whom were identified as heritage Spanish speakers using district criteria. The majority were born in the United States to parents of Mexican descent, although a small number of participants came from families originating in Central America. Participants exhibited a range of oral proficiencies in both languages, although Spanish language proficiency data were not collected in schools and were therefore not known to the researcher. In terms of English proficiency, the majority of children (approximately 75%) qualified for ESL services, while the rest either did not qualify or had already exited from such services. Therefore, children in this study ranged from beginning to advanced English speakers.

Three children from the original sample (*N* = 56) did not adequately complete the ONR in Spanish, and two other children did not complete it in English. Therefore, only 51 children completed assessments in both languages. Descriptive data for all 56 participants is reported (see Table 1), but sample sizes for each of the analyses vary accordingly.

Oral Narrative Retell Procedure

ONRs using two Mercer Mayer *Frog* wordless picture books (Mayer, 1969, 1974) and the Strong Narrative Retell Assessment Procedure (Strong, 1998) were conducted with children from February through May 2013. All assessments were given either by the principal investigator or a trained graduate student. The principal investigator is a Latina highly fluent Spanish-English bilingual who has experience teaching in a dual language elementary school, and has conducted previous research with Spanish-speaking populations. The graduate student was not bilingual and only administered English assessments.

Table 1
Descriptive Statistics

Variable	Sample (*N* = 56)
Grade	
First	23
Second	33
Gender	
Male	29
Female	27
Schooling context	
English only	30
Dual language	26

The Strong elicitation procedure calls for the assessor to give the child a general overview of the content of the book as well as instructions about how and when to turn pages while listening. Children were also told in advance that they would be asked to retell the story in their own words and that they would not be able to look at the book while doing so. Children listened to the story using headphones while the assessor sitting across from the child and worked on a computer. This naïve listener condition (Strong, 1998) is generally expected to lead to more detailed retelling because of the lack of shared knowledge between researcher and child. After listening, each child was given the option to review the pictures in the book before giving it to the assessor to put away. About half the children took the option to do so. The audiorecorder was turned on, and children were prompted to begin their retell. The assessor remained silent throughout, intervening only after pauses of more than 3 s. After 3 s, the assessor gave a general prompt such as "Tell me more" (*Dime más*) or "Anything else?" (*¿Algo más?;* Miller, Andriacchi, & Nockerts, 2011). Once the child had retold most of the story or paused for longer than 5 s, the researcher asked, "Is that all you remember?" (*¿Es todo lo que recuerdas?*).

All children were assessed in a quiet space in the school. Each was assessed twice: first in Spanish (by the principal investigator) and approximately a week later in English (by the principal investigator or a graduate research assistant). A counterbalanced design was employed, such that half of the children were randomly chosen to hear *Frog Goes to Dinner* in Spanish and *Frog, Where Are You?* in English. The opposite was true for the other half of children. The frog books are designed to be similar in terms of length, sentence complexity, and cohesion (Greenhalgh & Strong, 2001), and have been used across languages with Spanish speakers in other studies (Bedore et al., 2010; Simon-Cereijido & Gutiérrez-Clellen, 2009). Each assessment session lasted approximately 20 min, and only the target language for that session was spoken by the assessor for the entire session.

Data Analysis

All narratives were audiorecorded and transcribed either by the principal investigator or trained and experienced university students who were highly proficient in the target language. The Systematic Analysis of Language Transcripts (SALT) computer program and its accompanying conventions were used for all segmenting and coding (Miller, 2012; Miller et al., 2011). All measures in this study were calculated using only complete and intelligible utterances. In addition, maze behaviors such as repetitions and reformulations were omitted and therefore did not count in any of the measures. There were few instances of code switching in the data set, but those instances were also considered maze behavior and were not counted in final measures.

In line with SALT conventions, transcripts were segmented into C-units using Loban's (1976) rules; a C-unit includes a main clause and all of its subordinate clauses. For example, the sentence *The frog jumped before it landed in the water* would be considered one C-unit, since *before* is a subordinating conjunction. Segmenting of C-units only differed in the case of coordinated clauses with omitted subjects in the second main clause, as recommended in the literature on ONRs with Spanish-speaking children (Gutiérrez-Clellen & Hofstetter, 1994; Miller et al., 2006). In other words, utterances that contain a succession of coordinating verbs (*and, but, or,* etc.) without repeating the subject were segmented into separate C-units, even if they would be considered one utterance using standard segmenting conventions. For example, the following would be considered one utterance using standard C-unit segmenting conventions: *The frog jumped and landed in the water.* However, using modified C-units, as I did in the present study, it would be segmented as two utterances: *The frog jumped/and landed in the water.* This modified C-unit segmenting makes sense when analyzing transcripts in studies with bilingual Spanish speakers because of the pronoun-drop nature of Spanish; the subject pronoun is not necessary since it is encoded in the verb that follows.

Modified C-unit segmenting results in a greater overall number of utterances than might be the case with standard C-unit segmenting, but it protects against overinflating the grammatical complexity children are using. To maintain consistency and comparability across languages, modified C-unit segmenting was used in the present study to segment all transcripts in both languages. This same procedure was used by Miller et al. (2006) to ensure that participants' narrative abilities in both languages were calculated equitably (Miller et al., 2011).

NDW was the microlevel measure analyzed within the lexical domain (Miller et al., 2006). It is a total count of the unique, uninflected lexemes used in each complete retell. It is helpful as a cross-linguistic measure because it accounts for the very different morphological systems in English and Spanish and allows for direct comparison across languages (Simon-Cereijido & Gutiérrez-Clellen, 2009). NDW has been shown to be a developmentally sensitive and robust indicator of a child's vocabulary and is widely used in research on oral narrative production among bilingual children (Heilmann, Miller, Nockerts, & Dunaway, 2010). It has also been found to be positively related to reading achievement in Spanish-speaking children (Miller et al., 2006; Rojas & Iglesias, 2013). NDW is analyzed automatically by the SALT program.

Because it is based on breadth of vocabulary rather than depth, NDW is useful as a measure of general language ability. It is not meaningful as a measure of academic language or facility with complex language (Greenhalgh & Strong, 2001). In this study, therefore, two measures of grammatical complexity were analyzed. Grammatical analysis of ONRs is important because syntactic knowledge plays a role in reading comprehension, especially with regard to children's ability to synthesize information (Gutiérrez-Clellen, 1998; Verhoeven, 2011). The first grammatical component analyzed was mean length of utterance at the word (MLUw) level, widely considered a reliable measure of grammar for this population of learners (Castilla, Restrepo, & Perez-Leroux, 2009; Rojas & Iglesias, 2013). MLUw is analyzed automatically by the SALT program.

The second measure of grammatical complexity analyzed was the SI, which is the ratio of the total number of clauses (independent and subordinate) in a retell divided by the total number of utterances. SI has been used to analyze the ONRs of both monolingual and bilingual children (Gutiérrez-Clellen & Hofstetter, 1994; Heilmann, Miller, Nockerts, & Dunaway, 2010), and is considered to be a valid indicator of grammatical proficiency. Gutiérrez-Clellen and Hofstetter (1994) argued that the use of such an index to analyze the language production of young children "may reveal important information about their facility with school literacy activities that rely on the construction of spoken and written narratives" (p. 646). Research also suggests that the SI can be used to quantify complex grammar across the two languages of bilingual children (Miller et al., 2011). The SI is not automatically analyzed by SALT and was therefore hand-coded by the principal investigator in this study. It was then entered into the SALT transcript. Frequencies, means, and standard deviations were then calculated using SALT.

The final measure analyzed was the NSS (Heilmann, Miller, Nockerts, & Dunaway, 2010) to assess children's use of discourse. The NSS includes basic story grammar elements, but also incorporates sophisticated narrative elements like cohesion, mental states, and the use of clear antecedents and referents. It is an overall index of a child's ability to produce a coherent and detailed narrative. The NSS consists of seven categories, each of which is scored holistically on a scale of 1 (*minimal/immature*) to 5 (*proficient*) for a possible 35 points total (Heilmann, Miller, Nockerts, & Dunaway, 2010). Table 2 highlights key criteria used to assess each of the elements.

Reliability

To achieve interrater reliability, 20% of transcripts in each language (10 in Spanish, 9 in English) were randomly selected to be transcribed, segmented, and coded by both the researcher and a trained university student. In English, word-to-word match transcription was 95%, point-to-point agreement on identification of modified C-units was 98%, and the identification of subordinate clauses was 84%. In Spanish, word-to word match transcription was 92% and point-to-point identification of modified C-units was 93%. The researcher and assistant discussed points of disagreement to inform the transcription of the rest of the recorded retells.

Following Heilmann, Miller, and Nockerts (2010), Krippendorff's alpha (Krippendorff, 1980) was used to calculate interrater reliability for NSS, as it accounts for differences in both agreement and degree of difference on ordinal data between transcribers. Alpha values were calculated using the total NSS scores that were calculated by the two independent transcribers ($\alpha = .86$ for NSS). This was above Krippendorff's benchmark of .80 for adequate agreement.

Results

Table 3 shows descriptive statistics for the lexical, grammatical, and discourse components based on ONR scores in each language for all participating children as well as by instructional context.

Overall, the means for MLUw, SI, and NSS were similar across languages, with slightly higher scores in English in most areas; however, mean NDW scores were considerably higher in English than in Spanish (see Table 3). Means for dual language and English-only groups were calculated separately to ascertain whether scores differed significantly based on instructional context. As can be seen in Table 3, differences were observed between the two groups on several individual components, generally such that children in dual language immersion had higher scores than children in English-only classrooms. Independent groups t tests indicated that differences between the two groups were significant for Spanish NDW (SpNDW; $p < .01$), Spanish NSS (SpNSS; $p < .05$), English MLUw (EngMLUw; $p < .001$), and English SI (EngSI; $p < .001$). For this reason, LOI was included as a main effect and moderator of other predictors in the hierarchical multiple regression analyses. Grade level was also entered as a possible predictor of NSS in each language. It was not found to be significant for SpNSS and was therefore not included as a covariate. Grade was found to be significant for EngNSS, so it was controlled for in the regression equation.

Table 2
Narrative Macrostructure Measure: Narrative Scoring Scheme (Heilmann, Miller, & Nockerts 2010)

Language sampling context	Story retell
Total possible score	35 points
Scoring criteria	Seven elements: each scored on a 1 (*immature*) to 5 (*proficient*) point scale, then summed; 0 points possible if child does not include element or provides incomprehensible response
Elements assessed	*Introduction:* introduces setting and characters (0–5 points) *Character development:* differentiates between main and secondary characters *Mental states:* uses vocabulary that expresses thoughts or emotions of characters *Referencing:* appropriately uses pronouns, referents, and antecedents *Conflict resolution:* identifies major conflicts and resolutions that are critical to the development of the plot *Cohesion:* applies logical sequencing and smooth transitions *Conclusion:* provides general conclusion to the story in addition to final event

Table 3
Means and Standard Deviations of Lexical, Grammatical, and Discourse Components by Language for all Participating Children

	NDW (*SD*)	MLUw (*SD*)	SI (*SD*)	NSS (*SD*)
Spanish ($n = 53$)	59.4 (25.2)	7.5 (1.3)	1.2 (0.17)	19.3 (7.5)
DL ($n = 26$)	68.6 (22.5)	7.7 (1.1)	1.3 (0.16)	21.5 (7.1)
EO ($n = 27$)	50.4 (24.9)	7.2 (1.4)	1.2 (0.18)	17.2 (7.3)
English ($n = 54$)	67.6 (29.2)	7.7 (1.2)	1.2 (0.15)	20.1 (6.7)
DL ($n = 25$)	74.6 (29.9)	8.3 (1.0)	1.3 (0.16)	21.5 (7.0)
EO ($n = 29$)	61.6 (27.7)	7.1 (0.98)	1.1 (0.12)	19.0 (6.4)

Note. ($N = 56$). NDW = number of different words; MLUw = mean length of utterance, word; SI = subordination index; NSS = Narrative Scoring Scheme; DL = dual language instructional context; EO = English only instructional context.

Correlations

To address the first research question, regarding which microlevel lexical and grammatical components of ONRs were significantly correlated within and across languages, Pearson product–moment correlations were calculated. Two-tailed correlations were run because they are more conservative than one-tailed tests, thus accounting for the limited sample size in this study. As with descriptive statistics, correlations are reported for all children (Table 4), as well as for each LOI subgroup separately (Tables 5 and 6). In all cases, preliminary analyses were conducted to ensure no violation of the assumptions of normality, linearity, and homoskedasticity.

For all children with complete data, in Spanish, the microlevel lexical component (NDW) was significantly and positively correlated with both grammatical components (SpNDW and SpMLUw, $r = .276$, $p < .05$; SpNDW and SpSI, $r = .306$, $p < .05$). In English, there were no significant correlations among microlevel domains. Cross-linguistically, vocabulary scores were significantly and strongly positively correlated with one another (SpNDW and EngNDW, $r = .657$, $p < .01$) whereas MLUw and SI scores were not (SpMLUw and EngMLUw, $r = -.048$, *ns;* SpSI and EngSI, $r = -.127$, *ns*). Across domains and languages, Spanish vocabulary was significantly and positively correlated with one of the English grammatical components (SpNDW and EngSI, $r = .319$, $p < .05$), and English vocabulary was significantly and positively correlated across languages with both components in the Spanish grammatical domain (EngNDW and SpMLUw, $r = .284$, $p < .05$; EngNDW and SpSI, $r = .383$, $p < .01$). Correlations among components were also run for each LOI subgroup separately. Because of the small sample in each subgroup, a Bonferroni adjustment was used, and as such a significance level of $p < .01$ was set. Results for children enrolled in English-only instruction are reported in Table 5.

Table 4
Correlations Within and Across Languages for Children With Complete Data

	2	3	4	5	6
1. SpNDW	.657**	.276*	.245	.306*	.319*
2. EngNDW		.284*	.189	.383**	.204
3. SpMLUw			−.048	.668**	−.258
4. EngMLUw				−.062	.632**
5. SpSI					−.127
6. EngSI					

Note. ($N = 51$). EngNDW = English number of different words; SpMLUw = Spanish mean length of utterance, word; EngMLUw = English MLUw; SpSI = Spanish subordination index; EngSI = English SI.
*$p < .05$ (two-tailed). **$p < .01$ (two-tailed).

For children in English-only instruction, in Spanish, the microlevel lexical component (NDW) was not significantly correlated with either of the grammatical components (MLUw or SI). However, the two different grammatical components were significantly and positively correlated (SpMLUw and SpSI, $r = .613$, $p < .01$). In English, vocabulary was significantly and positively correlated with the SI (EngNDW and EngSI, $r = .544$, $p < .01$). Cross-linguistically, vocabulary scores were significantly and positively correlated with one another (SpNDW and EngNDW, $r = .807$, $p < .01$) whereas MLUw and SI scores were not. Across domains and languages, there were no significant correlations between vocabulary and either component in the grammatical domain. Correlation results for children enrolled in dual language instruction are reported in Table 6.

For children in dual language instruction, the microlevel lexical component was not significantly correlated with either grammatical component in Spanish or English. However, the two different grammatical components were significantly and positively correlated within Spanish (SpMLUw and SpSI, $r = .741$, $p < .01$) and within English (EngMLUw and EngSI, $r = .596$, $p < .01$). Cross-linguistically, there were no significant correlations within either the vocabulary or the grammatical domain. Within the grammatical domain, Spanish MLUw was significantly and negatively correlated with EngSI (SpMLU and EngSI, $r = -.673$, $p < .01$) in this group. Across domains and languages, Spanish

Table 5

Correlations Within and Across Languages for Students in English-Only Instruction

	2	3	4	5	6
1. SpNDW	.807*	.293	.066	.455	.412
2. EngNDW		.071	.172	.292	.544*
3. SpMLUw			−.017	.613*	−.111
4. EngMLUw				−.118	.371
5. SpSI					.154
6. EngSI					

Note. (N = 26). Alpha level adjusted using a Bonferroni adjustment to account for the small sample size. EngNDW = English number of different words; SpMLUw = Spanish mean length of utterance, word; EngMLUw = English MLUw; SpSI = Spanish subordination index; EngSI = English SI.
$^*p < .01$ (two-tailed).

Table 6

Correlations Within and Across Languages for Students in Dual Language Instruction

	2	3	4	5	6
1. SpNDW	.443	.152	.027	.087	−.039
2. EngNDW		.501	.012	.465	−.206
3. SpMLUw			−.331	.741*	−.673*
4. EngMLUw				−.151	.596*
5. SpSI					−.496
6. EngSI					

Note. (N = 25). Alpha level adjusted using a Bonferroni adjustment to account for the small sample size. EngNDW = English number of different words; SpMLUw = Spanish mean length of utterance, word; EngMLUw = English MLUw; SpSI = Spanish subordination index; EngSI = English SI.
$^*p < .01$ (two-tailed).

Table 7

Regression Models Predicting Spanish NSS Scores

Predictors	R^2	Adjusted R^2	R^2 change	F change	p value
Model 1: SpNDW	.733	.728	.733	134	< .001
Model 2: +SpMLUw, SpSI	.738	.721	.005	0.415	.662
Model 3: +EngNDW	.739	.717	.002	0.268	.607
Model 4: +EngMLUw, EngSI	.747	.712	.008	0.658	.523
Model 5: +LOI	.754	.714	.008	1.35	.253
Model 6: +LOI × SpNDW, LOI × SpMLUw, LOI × SpSI, LOI × EngNDW, LOI × EngMLU, LOI × EngSI	.811	.744	.056	1.84	.118

Note. (N = 51). NSS = Narrative Scoring Scheme; SpNDW = Spanish number of different words; SpMLUw = Spanish mean length of utterance, word; SpSI = Spanish subordination index; EngNDW = English NDW; EngMLUw = English MLUw; EngSI = English SI; LOI = language of instruction.

vocabulary was not significantly correlated with either grammatical component, nor was English vocabulary significantly correlated across languages with either component in the Spanish grammatical domain.

When patterns of correlation between the overall sample (Table 4), English-only students (Table 5), and dual language students (Table 6) were compared, differences emerged, indicating once again that LOI needed to be included as a main effect and moderator of other predictors in subsequent hierarchical multiple regression analyses. Results of such analyses are reported in the next section.

Regression Analyses

To address the second research question, regarding how well lexical and grammatical microlevel components predicted a macrolevel narrative discourse score within and across languages, two series of hierarchical multiple regressions were performed using the Spanish and English narrative scores as the dependent variable. Preliminary analyses were conducted to ensure that assumptions of normality, collinearity, and homoskedasticity were not violated. To avoid multicollinearity due to the inclusion of interaction terms, continuous predictors were mean-centered and LOI was effect-coded such that −1 represented students in English-only classrooms and 1 represented students in dual language immersion classrooms.

The first series of hierarchical multiple regression models used Spanish NSS scores as the dependent variable. The first two models in this series investigated the predictive value of each of the microlevel domains (NDW for lexical, MLUw and SI for grammatical) in the same language. English lexical and grammatical components were added in the third and fourth models. LOI, indicating enrollment in either dual language or English-only instruction, was entered in the fifth model, and interactions between LOI and each of the microlevel predictors were entered in the sixth model. Taking this sequential additive approach made it possible to first identify the variance

Table 8
Regression Models Predicting English NSS Scores[a]

Predictors	R^2	Adjusted R^2	R^2 change	F change	p value
Model 1: EngNDW	.767	.758	.520	107.19	< 0.001
Model 2: +EngMLUw, EngSI	.800	.783	.033	3.79	< 0.05
Model 3: +SpNDW	.826	.807	.026	6.75	< 0.01
Model 4: +SpMLUw, SpSI	.847	.823	.021	2.99	0.061
Model 5: +LOI	.847	.818	.000	.000	1.00
Model 6: +LOI × EngNdw, LOI × EngMLu, LOI × EngSI, LOI × SpNDW, LOI × SpMLU, LOI × SpSI	.857	.802	.010	.411	0.867

Note. (N = 51). NSS = Narrative Scoring Scheme; EngNDW = English number of different words; EngMLUw = English mean length of utterance, word; EngSI = English subordination index; SpNDW = Spanish NDW; SpMLUw = Spanish MLUw; SpSI = Spanish SI; LOI = language of instruction.
[a]Grade was controlled for in this model.

accounted for by components in the same language, then to identify any unique variance that was contributed by English microlevel components, and finally to understand the extent to which LOI accounted for unique variance beyond microlevel components and moderated the relations between microlevel components and Spanish NSS. As can be seen in Table 7, the total variance accounted for by the first set of regression equations was 73.3%, $F(13, 37) = 12.20$, $p < .001$. Spanish lexical diversity was the only significant predictor of Spanish NSS scores, accounting for 73% of the overall variance ($\beta = .856$, $p < .001$). Adding Spanish grammatical components did not add significantly to the explanatory power of the model, nor did adding English microlevel components. LOI and its interactions also did not contribute significantly, suggesting that the predictive value of microlevel components was not different for children in dual language and English-only settings.

The same model-building procedure was used to investigate the microlevel predictors of English NSS. That is, English microlevel lexical and grammatical domains were entered first in two models, followed by Spanish components, LOI, and finally interactions between LOI and Spanish and English microlevel predictors. Results can be seen in Table 8.

The total variance accounted for by this set of regression equations was 81.9%, $F(13, 37) = 16.6$, $p < .001$. As was the case with Spanish NSS, English lexical diversity was the most significant predictor of English NSS scores, accounting for 75% of the overall variance. However, unlike the case for Spanish NSS, adding English grammatical components to the final model also significantly explained variance. Cohen's f^2 was calculated to estimate effect size of this addition. According to Cohen (1988), effect sizes that are considered small, medium, and large are .02, .15, and .35 respectively. The 4% change in variance explained by the addition of grammatical components was significant, $F(3, 47) = 60.10$, $p < .05$, $f^2 = .12$, exhibiting a moderate effect size and demonstrating that grammar performance was predictive of children's NSS scores within English beyond the variance explained by NDW. In addition, the contribution of Spanish NDW was significant when added in the third model ($\beta = .222$, $p < .01$). The 3% change in variance was significant, $F(4, 46) = 52.14$, $p < .01$, $f^2 = .10$, again showing moderate effect size. LOI and its interactions were not significant, suggesting that the predictive value of microlevel components was not different for children in dual language and English-only settings.

Discussion

This study set out to investigate relations among the lexical, grammatical, and discourse domains of ONRs within and across the languages of first- and second-grade Spanish-speaking second language learners. Two specific relations were of interest: (a) associations within and across microlevel lexical and grammatical domains, and (b) the predictive validity of microlevel domains on a macrolevel discourse score. This dual focus highlights the critical role that oral language skills play in reading comprehension, especially for children learning a second language (Droop & Verhoeven, 2003; Gottardo, 2002; Gottardo & Mueller, 2009; Verhoeven, 2000). Correlations were run to examine the first relation, and regression analyses were run to examine the second.

In terms of microlevel correlations, it was hypothesized based on previous research that there would be significant correlations among domains within each language but not across languages.

The specific relations that could be expected were hard to predict because of inconsistent results in previous ONR and non-ONR studies with similar populations of children. Some of these inconsistent findings were the result of the use of different and noncomparable assessments across studies, or the use of noncomparable assessments across languages within studies (Simon-Cereijido & Gutiérrez-Clellen, 2009). Given such mixed results, it was not surprising that findings from the current study both confirm and contradict existing research. On one side, findings about within-Spanish correlations largely corroborated existing research (Gottardo, 2002; Gottardo & Mueller, 2009; Manis et al., 2004). Like all previous studies examining the relation between the Spanish lexical domain and the Spanish grammatical domain, results from the present study indicate significant positive correlations between NDW and both MLUw and SI for the overall sample. In contrast, however, findings about cross-linguistic lexical correlations contradict much of the extant research while confirming findings from a smaller number of studies. Multiple studies—using both the ONR methodology and standardized assessments—have reported nonsignificant correlations for vocabulary across languages, while the present results indicate a high level of significance ($p < .01$). The previous study that reported nonsignificance used the ONR methodology (Bedore et al., 2010), whereas the two studies that found significance used the Woodcock Language Proficiency Battery to measure vocabulary (Lindsey et al., 2003; Proctor et al., 2006). This suggests that it is not the assessment tool per se that matters, but that there are likely other factors at play, such as age and schooling context.

Findings from the present study also add to the mixed picture with regard to cross-linguistic, cross-domain correlations. Two studies using the ONR methodology with slightly younger children found nonsignificant correlations between NDW and MLUw across languages, whereas in the current sample, the correlation in at least one direction (EngNDW and SpMLUw) was significant. Further, there was a significant positive relation between NDW and SI—a grammatical measure that the other two studies did not calculate—across languages as well. In the current study, therefore, microlevel lexical to grammatical cross-linguistic correlations in the overall sample were quite robust, in contrast to the limited existing literature using the same methodology. There are at least three possible reasons for discrepancies between current findings and those from previous research. First, participants in the other two ONR studies reported in this article were in preschool or kindergarten, whereas children in the current study were nearing the end of first or second grade. Given that the early school years are pivotal in terms of language development for second language learners, it should not be surprising that even children a year or two older may exhibit different patterns of language development.

Second, although children's patterns of language use were described in the two other studies, it was not indicated in either whether children were enrolled in home language literacy instruction, leaving one to assume that they received instruction in English only. In the current study, nearly half (26/56) of the participants were enrolled in dual language education and had therefore received 2–3 years of Spanish reading instruction. Schooling context is an important consideration and may have played a role in patterns of cross-linguistic correlations found in this study. In fact, participants in this study who were enrolled in dual language education had better developed microlevel and macrolevel skills in both languages than children in English-only instruction.

A third possible reason for the discrepancy in microlevel findings may have been the use of the SI as a grammatical measure in the current study. Although MLUw is commonly considered a proxy for grammatical complexity, it only reflects the average number of words in each utterance, regardless of whether the syntactic structure of the utterance is complex or not. Therefore, SI was included to account specifically for complexity beyond length. The mean SI was not high for this sample in either language—1.2 in both Spanish and English—so it is notable that it was significantly correlated with lexical components within Spanish and across languages for the overall sample, as well as being a significant predictor of NSS scores within English.

In terms of microlevel predictors of the macrolevel discourse score, it was hypothesized that vocabulary and grammatical components would be significant predictors of NSS scores within both languages but not cross-linguistically. Findings were more nuanced than was hypothesized: Vocabulary was the only significant predictor of overall narrative performance within both languages, as well as cross-linguistically from Spanish to English; English grammar contributed significant variance to the regression equation for English narrative performance only. The findings around vocabulary as a predictor of narrative performance confirmed much of the existing research on the important role that vocabulary plays in language and literacy development in the early school years, including that which has used both standardized assessments and ONR methodology (Scarborough, 2001; Verhoeven, 2011). The finding that SI as a measure of grammatical complexity was more closely related to both microlevel and macrolevel components than was MLUw was somewhat more novel, since the two ONR studies that were previously conducted to investigate the role of grammar on oral narrative proficiency did not include an index of subordination (Bedore et al., 2010; Simon-Cereijido & Gutiérrez-Clellen, 2009). This suggests that more research using the SI as a measure of grammatical complexity is warranted.

The current study therefore adds to the existing knowledge base on vocabulary among Spanish-speaking children in three important ways. First, only two ONR studies have previously investigated the predictive value of vocabulary breadth on NSS scores, and both were conducted with English monolinguals (Heilmann, Miller, Nockerts, & Dunaway, 2010; Terry et al., 2013). In the current study, the same predictive pattern was found from vocabulary to discourse in English as the two previous studies, extending it to a second language learner population. Second, findings showed a similar highly significant predictive relation between Spanish vocabulary and Spanish discourse for these children, suggesting that a parallel relation may exist between microlevel and macrolevel domains in both languages. Third, vocabulary cross-linguistically predicted discourse scores from Spanish to English when added to the regression equation after English components, suggesting that L1 vocabulary is another critical predictor of L2 narrative skill. This is one of the first studies to report such a finding using the ONR assessment method, confirming what has been found with other tools and expanding understanding of the various ways in which lexical skills can be measured and assessed.

The role of grammatical development in oral narrative retelling was less clear from the results of this study. There were microlevel correlations between vocabulary and grammatical components both within and across languages, but grammatical performance significantly predicted discourse scores only within English. Within Spanish, there were no grammatical predictors of discourse performance. The reason for this is unknown, since children's mean scores in MLUw and the SI were similar across languages. Nonetheless, it confirms previous research that has found grammar to be a less robust predictor of narrative discourse than vocabulary (Droop & Verhoeven, 2003; Gottardo & Mueller, 2009; Lindsey et al., 2003; Manis et al., 2004).

Findings from this study confirmed much of what has been done in previous research even though participants were enrolled in programs with different languages of instruction (Gottardo, 2002; Gottardo & Mueller, 2009; Manis et al., 2004; Miller et al., 2006). Second language learners from two instructional contexts performed similarly on the ONR task, with both groups generally performing better in English than in Spanish, and dual language children performing slightly better than English-only students on all components in both languages.

Finally, it is worth highlighting that while LOI was not found to be a significant predictor or moderator of other predictors in hierarchical multiple regression analyses, there were some differences in the performance of English-only and dual language students on retells in the two languages. For example, among children in dual language classrooms, there was a negative cross-linguistic correlation between the two grammatical components. Such differences were not found to be statistically meaningful, but they may well be practically significant, and further research that analyzes such differences is warranted.

Implications

Findings from this study confirm the critical roles that vocabulary and grammatical complexity play in the development of narrative discourse skill for young second language learners. They also reiterate that ONRs provide a window into linguistic development, as well as the development of coherent, literate stories. ONR data are especially informative when analyzed in two languages because of the interdependence of skills found across microlevel domains. In concert with previous research in this area, findings from the current study have important implications for both early literacy instruction and future research.

With regard to instruction, others have noted the value of a focus on vocabulary for second language learners (Droop & Verhoeven, 2003; Verhoeven, 2011), and indeed, findings from the current study confirm that the number of words available to second language learners in two languages may affect their facility with narrative production. Therefore, attention should be paid to building second language learners' breadth of vocabulary in both languages, in addition to supporting children as they learn targeted words in depth (August et al., 2005; Carlo et al., 2004). In addition, it was found that vocabulary and grammar are related within the L1 and across languages, but more importantly, that grammatical complexity is a possible predictor of oral discourse skills, at least within English. An instructional implication, then, is that facilitating complex syntax should be a goal of early literacy instruction for second language learners. This would require educators to think deeply about how such complexity can be taught and supported within both dual language and English-only programs.

In terms of future research, findings from this study suggest that patterns of development may be similar in each of a child's two languages, even if they do not have access to instruction in their home language. A limitation of the current study is that the sample size was too small to draw conclusions about the development of narrative skill for children enrolled in two different instructional contexts, although there are hints that such differences may exist. Nonetheless, regression equations showed that enrollment in dual language or English-only literacy instruction did not moderate children's performance in either language, and the population studied mirrors that of the general Spanish-speaking population in the United States, given the variety of bilingual programs in which such children are enrolled (Baker, 2011; García & Kleifgen, 2010). A second limitation was the lack of Spanish language proficiency information available about participating children. Future research should further investigate similarities and differences in the oral language of primary grade children in different instructional contexts using a culturally relevant and holistic assessment like oral narrative retelling. In addition, given the finding that the SI was a significant predictor of NSS scores within English but not within Spanish for the overall sample, it would be worth further investigating how this grammatical skill develops in two languages, as well as identifying possible ways to facilitate the development of syntactically complex language in this population.

In summary, findings from the present study contribute to the field's understanding of within- and cross-linguistic relations in the ONRs of Spanish-speaking children, and begin to untangle how second language learners process and produce retells in their first and second languages. ONRs are a nuanced and powerful tool to use in assessing the growing population of young second language learners in American schools, and offer a more complete picture of such children's potential literacy comprehension strengths and challenges than other commonly used tools. Perhaps more importantly, research on oral narrative development can elucidate ways to support children as they develop language and literacy skills in two languages.

References

Allen, M. M., Ukrainetz, T. A., & Carswell, A. L. (2012). The narrative language performance of three types of at-risk first-grade readers. *Language, Speech, and Hearing Services in Schools, 43*, 205–221. doi:10.1044/0161-1461 (2011/11-0024)

August, D., Carlo, M., Dressler, C., & Snow, C. E. (2005). The critical role of vocabulary development for English language learners. *Learning Disabilities Research and Practice, 20*, 50–57. doi:10.1111/j.1540-5826.2005.00120.x

Baker, C. (2011). *Foundations of bilingual education and bilingualism* (5th ed.). Clevedon, UK: Multilingual Matters.

Bedore, L. M., Peña, E. D., Gillam, R. B., & Tsung-Han, H. (2010). Language sample measures and language ability in Spanish-English bilingual kindergartners. *Journal of Communication Disorders, 43*, 498–510. doi:10.1016/j.jcomdis.2010.05.002

Bialystok, E. (2001). *Bilingualism in development: Language, literacy and cognition*. Cambridge, UK: Cambridge University Press.

Bialystok, E., & Herman, J. (1999). Does bilingualism matter for early literacy? *Bilingualism: Language and Cognition, 2*, 35–44.

Carlisle, J. F., Beeman, M., Davis, L. H., & Spharim, G. (1999). Relationship of metalinguistic capabilities and reading achievement for children who are becoming bilingual. *Applied Psycholinguistics, 20*, 459–478.

Carlo, M. S., August, D., McLaughlin, B., Snow, C. E., Dressler, C., Lippman, D. N., & White, C. E. (2004). Closing the gap: Addressing the vocabulary needs of English-language learners in bilingual and mainstream classrooms. *Reading Research Quarterly, 39*, 188–215. doi:10.1598/RRQ.39.2.3

Castilla, A. P., Restrepo, A., & Perez-Leroux, A. T. (2009). Individual differences and language interdependence: A study of sequential bilingual development in Spanish-English preschool children. *International Journal of Bilingual Education and Bilingualism, 12*, 565–580. doi:10.1080/13670050802357795

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Droop, M., & Verhoeven, L. T. (2003). Language proficiency and reading ability in first- and second-language learners. *Reading Research Quarterly, 38*, 78–103. doi:10.1598/RRQ.38.1.4

Durgunoglu, A. Y., Nagy, W. E., & Hancin-Bhatt, B. J. (1993). Cross-language transfer of phonological awareness. *Journal of Educational Psychology, 85*, 453–465. doi:10.1037/0022-0663.85.3.453

Fiestas, C. E., & Peña, E. D. (2004). Narrative discourse in bilingual children: Language and task effects. *Language, Speech, and Hearing Services in Schools, 35*, 155–168.

García, O., & Kleifgen, J. (2010). *Educating emergent bilinguals: Policies, programs, and practices for English language learners*. New York, NY: Teacher College Press.

Goodrich, J. M., Lonigan, C., & Farver, J. M. (2013). Do early literacy skills in children's first language promote development of skills in their second language? An experimental evaluation of transfer. *Journal of Educational Psychology, 105*, 414–426. doi:10.1037/a0031780

Gottardo, A. (2002). The relationship between language and reading skills in bilingual Spanish-English speakers. *Topics in Language Disorders, 22*, 46–70.

Gottardo, A., & Mueller, J. (2009). Are first- and second-language factors related in predicting second-language reading comprehension? A study of Spanish-speaking children acquiring English as a second language from first to second grade. *Journal of Educational Psychology, 101*, 330–344. doi:10.1037/a0014320

Greenhalgh, K. S., & Strong, C. J. (2001). Literate language features in spoken narratives of children with typical language and children with language impairments. *Language, Speech, and Hearing Services in Schools, 32*, 114–125.

Gutiérrez-Clellen, V. F. (1998). Syntactic skills of Spanish-speaking children with low school achievement. *Language, Speech, and Hearing Services in Schools, 29*, 207–215.

Gutiérrez-Clellen, V. F., & Hofstetter, R. (1994). Syntactic complexity in Spanish narratives: A developmental study. *Journal of Speech and Hearing Research, 37*, 645–654.

Hammer, C. S., & Miccio, A. W. (2006). Early language and reading development of bilingual preschoolers from low-income families. *Topics in Language Disorders, 26*, 322–337.

Heilmann, J., Miller, J. F., & Nockerts, A. (2010). Sensitivity of narrative organization measures using narrative retells produced by young school-age children. *Language Testing, 27*, 603–626. doi:10.1177/0265532209355669

Heilmann, J., Miller, J. F., Nockerts, A., & Dunaway, C. (2010). Properties of the Narrative Scoring Scheme using narrative retells in young school-age children. *American Journal of Speech-Language Pathology, 19*, 154–166. doi:10.1044/1058-0360(2009/08-0024)

Krippendorff, K. (1980). *Content analysis: An introduction to its methodology*. Thousand Oaks, CA: Sage.

Lindsey, K. A., Manis, F. R., & Bailey, C. E. (2003). Prediction of first-grade reading in Spanish-speaking English language learners. *Journal of Educational Psychology, 95*, 482–494. doi:10.1037/0022-0663.95.3.482

Loban, W. (1976). *Language development: Kindergarten through grade twelve* (Research report). Urbana, IL: National Council of Teachers of English.

Manis, F. R., Lindsey, K. A., & Bailey, C. E. (2004). Development of reading in grades K-2 in Spanish-speaking English language learners. *Learning Disabilities Research and Practice, 19*, 214–224. doi:10.1111/j.1540-5826.2004.00107.x

Mayer, M. (1969). *Frog, where are you?*. New York, NY: Dial Press.

Mayer, M. (1973). *Frog on his own*. New York, NY: Dial Press.

Mayer, M. (1974). *Frog goes to dinner*. New York, NY: Dial Press.

Mayer, M. (1975). *One frog too many*. New York, NY: Dial Press.

Miller, J. F. (2012). *Systematic analysis of language transcripts* [Computer software]. Madison: University of Wisconsin.

Miller, J. F., Andriacchi, K., & Nockerts, A. (2011). *Assessing language production using SALT software*. Middleton, WI: SALT Software.

Miller, J. F., Heilmann, J., Nockerts, A., Iglesias, A., Fabiano, L., & Francis, D. J. (2006). Oral language and reading in bilingual children. *Learning Disabilities Research and Practice, 21*, 30–43.

Muñoz, M. L., Gillam, R. B., Peña, E. D., & Gulley-Faehnle, A. (2003). Measures of language development in fictional narratives of Latino children. *Language, Speech, and Hearing Services in Schools, 34*, 332–342. doi:0161-1461/03/3404-0332

Oregon, D. O. E. (2013). *October 1 enrollment by ethnicity*. Retrieved from http://www.ode.state.or.us/sfda/reports/r0067Select2.asp

Proctor, C. P., August, D., Carlo, M., & Snow, C. E. (2006). The intriguing role of Spanish language vocabulary knowledge in predicting English reading comprehension. *Journal of Educational Psychology, 98*, 159–169. doi:10.1037/0022-0663.98.1.159

Proctor, C. P., August, D., Snow, C., & Barr, C. D. (2010). The interdependence continuum: A perspective on the nature of Spanish-English bilingual reading comprehension. *Bilingual Research Journal, 33*, 5–20. doi:10.1080/15235881003733209

Proctor, C. P., Carlo, M., August, D., & Snow, C. (2005). Native Spanish-speaking children reading in English: Toward a model of comprehension. *Journal of Educational Psychology, 97*, 246–256. doi:10.1037/0022-0663.97.2.246

Reese, E., Sparks, A., & Suggate, S. (2012). Assessing children's narratives. In E. Hoff (Ed.), *Research methods in child language: A practical guide* (pp. 133–148). Chichester, UK: Blackwell.

Rojas, R., & Iglesias, A. (2013). The language growth of Spanish-speaking English language learners. *Child Development, 84*, 630–646. doi:10.1111/j.1467-8624.2012.01871.x

Roth, F. P., Speece, D. L., & Cooper, D. H. (2002). A longitudinal analysis of the connection between oral language and early reading. *Journal of Educational Research, 95*, 259–272. doi:10.1080/00220670209596600

Roth, F. P., Speece, D. L., Cooper, D. H., & De La Paz, S. (1996). Unresolved mysteries: How do metalinguistic and narrative skills connect with early reading? *Journal of Special Education, 30*, 257–277. doi:10.1177/002246699603000303

Scarborough, H. S. (2001). Connecting early language and literacy to later reading dis(abilities): Evidence, theory, and practice. In S. B. Neuman & D. K. Dickinson (Eds.), *Handbook of early literacy research* (pp. 97–110). New York, NY: Guilford.

Simon-Cereijido, G., & Gutiérrez-Clellen, V. F. (2009). A cross-linguistic and bilingual evaluation of the interdependence between lexical and grammatical domains. *Applied Psycholinguistics, 30,* 315–337. doi:10.1017/S0142716409090134

Snow, C. E., Tabors, P. O., Nicholson, P. A., & Kurland, B. F. (1995). SHELL: Oral language and early literacy skills in kindergarten and first-grade children. *Journal of Research in Childhood Education, 10,* 37–47. doi:10.1080/02568549509594686

Speece, D. L., Roth, F. P., Cooper, D. H., & De La Paz, S. (1999). The relevance of oral language skills to early literacy: A multivariate analysis. *Applied Psycholinguistics, 20,* 167–190. doi:10.1017/S0142716499002015

Strong, C. J. (1998). *Strong Narrative Assessment Procedure.* Eau Claire, WI: Thinking.

Terry, N. P., Mills, M. T., Bingham, G. E., Mansour, S., & Marencin, N. (2013). Oral narrative performance of African American prekindergartners who speak nonmainstream American English. *Language, Speech, and Hearing Services in Schools, 44,* 291–305. doi:10.1044/0161-1461(2013/12-0037)

Verhoeven, L. T. (2000). Components in early second language reading and spelling. *Scientific Studies of Reading, 4,* 313–330. doi:10.1207/S1532799XSSR0404_4

Verhoeven, L. T. (2011). Second language reading acquisition. In M. L. Kamil, P. D. Pearson, E. B. Moje, & P. P. Afflerbach (Eds.), *Handbook of reading research* (Vol. 4, pp. 661–683). New York, NY: Routledge.

Child Development, September/October 2015, Volume 86, Number 5, Pages 1434–1448

Knowing When to Be "Rational": Flexible Economic Decision Making and Executive Function in Preschool Children

Wendy S. C. Lee and Stephanie M. Carlson
University of Minnesota

Failure to delay gratification may not indicate poor control or irrationality, but might be an adaptive response. Two studies investigated 3.5- and 4.5-year-old children's ability to adapt their delay and saving behavior when their preference (e.g., to delay or not delay) became nonadaptive. In Study 1 ($N = 140$), children's delay preference was associated with a risk of losing rewards. In Study 2 ($N = 142$), children's saving preference was associated with an inability to play an attractive game. Whereas baseline delaying and saving preferences were unrelated to a standardized executive function measure, children who switched to their nonpreferred choice scored higher, suggesting flexibility of decision-making may be a more meaningful dependent variable than baseline performance in developmental research on self-control.

The ability to resist the temptation of distractions and forgo immediate pleasure in pursuit of long-term goals has relevance for many domains of functioning, including health (e.g., addiction, nutrition, exercise), finances (e.g., spending, saving, investing), relationships (e.g., marriage, parenting), and educational and career achievement (e.g., studying, working). Delay of gratification can be defined as the postponing of immediate gratification to attain a delayed more valuable reward (e.g., Mischel, Shoda, & Rodriguez, 1989), and the underlying self-control processes have roots in early childhood. In their classic laboratory paradigm, Mischel and colleagues measured how long preschool children were able to wait when given the choice between having one marshmallow now and waiting for two marshmallows later (e.g., Mischel, 1974; Mischel et al., 1989). They discovered that individual differences in wait times and delay behavior during early childhood predicted a range of developmental outcomes into adolescence and adulthood, including academic competence and SAT scores, self-regulation, effective coping with stress and frustration, and positive peer relations (e.g., Ayduk et al., 2000; Mischel, Shoda, & Peake, 1988; Shoda, Mischel, & Peake, 1990). Remarkably, early delay behavior even predicted inhibitory control 40 years later (Casey et al., 2011), suggesting long-term stable individual differences in delay of gratification (see Mischel et al., 2011, for review).

Delay of gratification has significant social and economic benefits for both individuals and society, yet it appears to be a consistent struggle not only for children but also for many adults. If rational behavior and action maximizes utility, how do we explain failures to delay gratification? The underlying assumption is that the rational economic choice is always to delay or wait for the larger reward, and therefore, failures to do so are interpreted as deficient self-control. However, this is a misguided assumption. Whether a smaller immediate reward or a larger later reward is more adaptive is determined by the intersection of proximal and distal factors that contextualize the problem in important ways. For example, if we do not trust we will receive the later reward, or we have greater immediate needs, the wiser choice would be the smaller reward. In some contexts, not delaying is the rational response, and consequently, delay behavior may not always reflect self-control capacity.

Contextual Influences on Delay Behavior

There is renewed interest in understanding the rationality behind delay behavior, or lack of delay

---

We are grateful to the children and families who participated in this project, and thank our research assistants and staff, in particular Hannah Saunders, for their assistance in conducting this investigation. Funding was provided by the Social Sciences and Humanities Research Council of Canada (Wendy S. C. Lee) and by R01HD1495 (Stephanie M. Carlson). Stephanie M. Carlson is entitled to royalties from the sale of the Minnesota Executive Function Scale by Reflection Sciences, LLC.

Correspondence concerning this article should be addressed to Wendy S. C. Lee, Institute of Child Development, University of Minnesota, 51 East River Parkway, Minneapolis, MN 55455. Electronic mail may be sent to leex3325@umn.edu.

© 2015 The Authors
Child Development © 2015 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2015/8605-0009
DOI: 10.1111/cdev.12401

behavior. For example, cultural customs and values can powerfully shape behavior toward delaying or not delaying (Carlson & Zelazo, 2011; Mischel, 1958). In their findings on Hawaiian-American children, Gallimore, Weiss, and Finney (1974) discuss how the cultural importance of immediate consumption during social gatherings for building relationships, affiliation, and social credit would encourage not delaying as an adaptive behavior. More recent studies have focused on expectations for receiving rewards. In an experiment by Kidd, Palmeri, and Aslin (2013), 4-year-old children were randomly assigned to a situation in which promised desirable art supplies were either delivered or not delivered by the experimenter. Children who found the adult to be unreliable waited significantly less time during a subsequent delay task than children who experienced reliability, replicating earlier findings with older children by Mahrer (1956). Expectations can also be influenced by race (Strickland, 1972) or by simply pairing delay choices with untrustworthy faces and character vignettes (Michaelson, de la Vega, Chatham, & Munakata, 2013). Statistical information in the environment can also shape our expectations. When there is uncertainty about when a future reward will arrive, how long we persist or do not persist should depend on our experience of previous delays periods. McGuire and Kable (2012) showed that when presented with choices between an immediate small reward and waiting for an unknown amount of time for a larger later reward, over time, adults calibrated their persistence based on the length of delays they experienced.

Taken together, these studies provide collective evidence that delay of gratification and its associated paradigms are sensitive to contextual factors. Sometimes *not* delaying is guided by rational economic considerations, and not by poor self-control.

Delay of Gratification and Saving

This context dependency is not unique to delay of gratification, but is characteristic of other related behaviors such as saving. For example, although it is often better to save, it may be rational to spend your current resources if you do not foresee a stable long-term future. In discussions of delay of gratification, saving is often not clearly distinguished from delaying, and sometimes the ability to delay gratification is seen as a precursor to saving (e.g., Ainslie, 1975; Wärneryd, 1999). Wärneryd (1999) describes saving as the postponing of some consumption for future enjoyment, and while delay of gratification and saving both require inhibiting impulses toward immediate rewards and both unfold across time, some argue that they retain important differences (Sonuga-Barke & Webley, 1993). Delay of gratification tests children's waiting ability, particularly in Mischel's paradigm, where waiting is driven by a desire for the larger reward, and waiting is the cost to attain this reward. Saving can be more future oriented such that even if the later reward is not larger, saving may be driven by future enjoyment or needs, and draws upon an understanding that present spending behaviors have future consequences (Metcalf & Atance, 2011; Otto, Schots, Westerman, & Webley, 2006).

There is little laboratory research on children's saving, although Sonuga-Barke and Webley (1993) and Otto et al. (2006) have examined saving in 4- to 12-year-old children using a simulated economy board game with tokens and opportunities to save and spend, revealing some improvements with age. Only one recent study has examined saving experimentally in preschoolers. Metcalf and Atance (2011) presented children with a chance to play a small marble chute game followed by a big marble chute game, but were only given three marbles. This required children to save marbles from the smaller game to play with the bigger game. No age differences were observed, although a subset of children who did not save the first time saved the second time when the task was repeated. To our knowledge, there has been no published research directly examining the relation between delay of gratification and saving in young children.

The Role of Executive Function

Although delay of gratification is often seen as measuring aspects of self-control, surprisingly it does not relate consistently to measures of executive function (EF). EF refers to the neurocognitive processes involved in the conscious control of emotions, thoughts, and behaviors, including working memory, inhibitory control, and set shifting (Carlson, Zelazo, & Faja, 2013; Miyake, Friedman, Emerson, Witzki, & Howerter, 2000). EF undergoes significant development over the preschool years, as reflected by age-related improvements on a wide range of EF tasks (Carlson, 2005; Garon, Bryson, & Smith, 2008).

Delay-of-gratification tasks often do not relate to or load on the same factors as typical EF tasks (e.g., Dalen, Sonuga-Barke, Hall, & Remington, 2004; Olson, Schilling, & Bates, 1999). Related delay tasks that require inhibition and waiting for a reward (e.g., not touching a gift or snack) also tend to load separately from EF measures such as Stroop tasks, and relate differently to developmental outcomes (e.g., Carlson & Moses, 2001; Carlson, White, & Davis-Unger, 2014; Murray & Kochanska, 2002; Smith-Donald, Raver, Hayes, & Richardson, 2007; Willoughby, Kupersmidt, Voegler-Lee, & Bryant, 2011). This lack of clear relation between delay and EF, while initially surprising, may be logical if delay behavior is influenced by context and delay tasks measure more than self-control. The *adaptation* of one's delay behavior, however, may be a clearer reflection of self-control. As noted above, there are some strong preferences and individual stability in delay behavior. When a situation calls for behavior that conflicts with one's desires or predispositions, delaying when one prefers immediate gratification, or acting on impulse when one would normally refrain, the recruitment of self-control processes may be especially important.

In particular, set shifting (i.e., switching between different mental sets, functions, or rules) may be associated with one's ability to adjust behavior across situations. In preschoolers, a common measure of EF is the Dimensional Change Card Sort (DCCS; Zelazo, 2006), which involves sorting bivalent cards (e.g., red rabbit, blue boat) with target cards (e.g., blue rabbit, red boat), by one set of rules (e.g., color), and then switching and sorting by an incompatible set of rules (e.g., shape). Younger children often perseverate and continue sorting by the first set of rules, despite knowing the new rules. Flexibly switching to the second set of rules improves with age and may involve an increased ability to reflect on and hierarchically integrate higher order rules (e.g., if it is the color game, follow these rules; if it is the shape game, follow these other rules; Zelazo, Müller, Frye, & Marcovitch, 2003). As EF matures, it can also facilitate greater attention to relevant contextual information, leading to reappraisal and reevaluation of different response options (Cunningham & Zelazo, 2007). We propose that the same processes involved in reappraisal and switching between incompatible rules might enable children to flexibly switch behaviors across environmental contexts (e.g., if context A, then delay; if context B, then do not delay).

The Current Studies

Previous studies have shown that delay behavior can be influenced by experimentally manipulating the context (e.g., expectations of rewards) and, hence, the adaptiveness of delaying. Nonetheless, many have relied primarily on showing group differences rather than individual change. Specifically, they do not address individual differences in the ability to respond adaptively to different contexts, especially when it conflicts with one's natural tendencies or desires. Furthermore, many have only examined change in one direction, toward not delaying. The goal of Study 1 was to examine whether individual differences in children's adaptation of their delay behavior in either direction—when placed in a situation that demands behavior opposite to their preferences—would be related to EF. This is in contrast to baseline delay behavior, which is often unrelated to EF tasks.

The goal of Study 2 was to replicate and strengthen findings by addressing the same research question using the same methodology, but with a different decision-making behavior: saving. Saving and delay of gratification share key similarities, such as inhibition toward immediate rewards and sensitivity to context, but may differ in such aspects as future-oriented thinking. Consistent results across the two studies would lend confidence that our findings are not specific to a paradigm, but rather generalize to a family of economic decision-making behaviors.

We designed a novel dependent variable to evaluate flexibility in delay and saving behavior. Using a within-subjects design, we assessed children's baseline preference, followed by a second assessment under new conditions that encouraged a switch in behavior. In Study 1, children were offered a series of choices between less treats immediately and more treats later in a delay choice paradigm (e.g., Mischel, 1966; Schwarz, Schrager, & Lyons, 1983; Thompson, Barresi, & Moore, 1997). In Study 2, children were offered a chance to play with a small marble chute game followed by a big marble chute game, but were given only three marbles, requiring them to save marbles from the smaller game (Metcalf & Atance, 2011).

We expected that individual differences in delay or saving flexibility would be related to various factors. First, we predicted that older children would show greater flexibility than younger children due to increasing EF with age (Carlson et al., 2013). Second, we predicted that there would be more switching in delayers and savers, where the adaptive behavior is shifting toward immediate rewards and spending, than in nondelayers and spenders, where the adaptive behavior is shifting toward waiting and saving. It is generally less taxing to be impulsive and indulge in immediate consumption, and thus switching in that direction should be less difficult. Third, we predicted that switching would be related to performance on a standardized set-shifting task, in keeping with our theory that the same EF processes that underlie DCCS performance also underlie adaptive decision making. In contrast, baseline delaying and saving were expected to be only weakly related to age and set shifting, given prior research (e.g., Beck, Schaefer, Pang, & Carlson, 2011; Hongwanishkul, Happaney, Lee, & Zelazo, 2005; Metcalf & Atance, 2011; Schwarz et al., 1983). In other words, although past research has not shown clear relations between measures of delay and EF, we propose that relations will emerge when the dependent measure of decision making is flexibility and adaptation, rather than baseline preferences.

Study 1

Children's ability to be flexible with their delay behavior was examined using the delay choice paradigm with food treats. Children's baseline preference for delaying was assessed and children were then administered the task a second time with modifications to the delivery of treats such that their preference at baseline (i.e., to delay or not delay) was associated with a risk of losing treats. This should create incentive for children to shift away from baseline choices that have become risky, and toward less preferred choices that are not risky.

Method

Participants

Children were recruited from April 2011 to June 2012 from a large U.S. Midwestern city using a university database of families that had expressed interest in research. Screening criteria included full-term birth and absence of serious medical conditions and developmental disorders. A total of 41 children were excluded because they showed no clear preference to delay or not delay (see also Bandura & Mischel, 1965). Excluded were thirteen 3.5-year-olds (10 females) and twenty-eight 4.5-year-olds (18 females). Significantly more 4.5-year-olds, $\chi^2(1, N = 181) = 7.75$, $p = .004$, and more females, $\chi^2(1, N = 181) = 6.88$ $p = .007$, were excluded. This may be related to the finding (see Results below) that older children and females were more likely to delay and, accordingly, more likely to delay four or five trials than younger children and males. Excluded children did not differ from the included sample on any other demographic characteristics. An additional 10 children failed to start or complete tasks due to fatigue, lack of interest, or noncompliance, and were excluded from analyses. The final sample consisted of 140 children, including seventy-nine 3.5-year-olds ($M = 42.42$ months, $SD = 0.68$, 35 females) and sixty-one 4.5-year-olds ($M = 54.41$ months, $SD = 0.67$, 28 females). The majority of children were White and non-Hispanic (84.3%), with the remaining children identified as White and Hispanic, African American, Asian, Native Hawaiian and Pacific Islander, or mixed ethnicity. Families were predominantly two-caregiver households (96.4%) and the majority of primary caregivers, mostly mothers, reported attaining a bachelor's degree or higher (84.3%). The median and mode household income was $75,000–$99,999.

Procedure and Measures

Families were invited to the lab where children were assessed individually in a child-friendly laboratory room during a videotaped 45-min session. Parents completed a demographics questionnaire. Children first completed the standard delay choice task (9 trials) and were categorized as baseline delayers (delayed 6–9 trials), baseline nondelayers (delayed 0–3 trials), or were excluded from analyses due to no clear preference (delayed 4 or 5 trials).

Children were administered a standardized receptive vocabulary test, as verbal ability likely plays a role in task comprehension and is often related to EF (Carlson et al., 2013). Then children were randomly assigned to either an experimental or a control condition and administered a modified version of the delay choice task that introduced a risk of losing treats. In the experimental condition, children were implicitly encouraged to choose their nonpreferred choice (i.e., delayers to not delay, nondelayers to delay) by associating risk with their preferred choice. A total of 77 children were assigned to the experimental condition, including forty-five 3.5-year-olds (20 females) and thirty-two 4.5-year-olds (17 females). In the control-risk condition, children were encouraged to choose their preferred choice (i.e., delayers to delay, nondelayers to not delay) by associating risk with their nonpreferred choice. This control-risk condition assessed the extent of spontaneous switching when there was no adaptive reason to do so. A total of 33 children were assigned to the control-risk condition, including seventeen 3.5-year-olds (7 females) and sixteen 4.5-year-olds (6 females). An additional control-no-risk group was included to replicate previous findings of test–retest reliability (Beck et al., 2011). The baseline task was repeated with the modified task stimuli but without risk. The 30 children in this group included seventeen 3.5-year-olds (8 females) and thirteen 4.5-year-olds (5 females). We expected that children's behavior in these two control groups would be largely consistent in the absence of a rationale for switching, thus ensuring that switching in the experimental group was due to the manipulation.

After the second delay choice task, children were again categorized as delayers (5–9 trials) or nondelayers (0–4 trials). Children who delayed 4 or 5 trials were not excluded here as the number of trials delayed did not reflect children's natural preference in the absence of any intervention, as it did during baseline. Instead, children likely were influenced by our experimental manipulation to shift their responding toward delaying 4 or 5 trials, as supported by our analyses below. Although it can be argued that these children showed no preference in the second delay task, it would still indicate a shift in behavior toward a more adaptive response as compared to baseline. If children's classification changed across the two tasks (e.g., delayer to a nondelayer), they were considered switchers. If children's classification did not change (e.g., remained a delayer), they were considered nonswitchers. All children then completed an independent measure of set-shifting ability.

Baseline delay choice. This task was adapted from Thompson et al. (1997). Children chose one of several treats to use in the game (i.e., small crackers, sugary cereal, raisins). They were then offered a series of choices between having one treat now, which was placed in a bowl and could be eaten immediately, and more treats later, which were placed in an opaque envelope to be returned at the end of the session. The side of presentation for the treats was counterbalanced. There were 9 trials in total balanced across three types of contingencies (1 vs. 2; 1 vs. 4; 1 vs. 6) presented in a fixed mixed order. The experimenter demonstrated both now and later choices for herself. For each test trial, children were presented with the treats and asked to make a choice (e.g., "Do you want to choose 1 now or 4 later?").

Modified delay choice. In the modified task, the same treats, number of trials, order, and contingencies were used. Two panels with clear plastic tubes were introduced and placed side by side in front of the child. On each panel, treats were deposited down a single tube that split into two separate tubes (an inverted Y). A hidden tab at the back controlled which tube the treats went down. In one panel, two bowls were placed under the two lower tubes to catch the treats. In this case, there was a 100% chance of receiving the treats. In the other panel, a bowl was placed under only one tube and the other tube led to an opaque sealed box. In this case, children were told that there was a chance the treats would get stuck in the box and not be delivered.

Each choice (now or later) was associated with a different panel. In the experimental condition, the risky panel was associated with the "later" choice for delayers and with the "now" choice for nondelayers. In the control-risk condition, the risky panel was associated with the "now" choice for delayers and the "later" choice for nondelayers. Children in the control-no-risk condition had two nonrisky panels. The experimenter demonstrated how both panels worked, and demonstrated both now and later choices for herself. During test trials, the risky panel was rigged such that treats always fell into the box and were lost. To mirror this behavior, the nonrisky panel was also rigged such that the treats consistently fell into the bowl on the right side. At the end, children were asked to identify which panel they liked best.

Peabody Picture Vocabulary Test–4 (PPVT–4). In this standardized test of receptive vocabulary (Dunn & Dunn, 2007), children were shown a booklet and asked to point to one of four pictures on each page, corresponding to a word spoken by the experimenter. Age-standardized scores were used in analyses.

Minnesota Executive Function Scale (MEFS™). This task was developed by Carlson (2012) and adapts the DCCS (Zelazo, 2006) for a wider age range. The MEFS has been shown to be a valid measure of EF for children aged 2–7 with good test–retest reliability (intraclass correlation = 0.93; Beck et al., 2011; Carlson, 2012), showing convergent validity with other EF measures, including the Blair–Willoughby battery and the Head–Toes–Knees–Shoulders task, *r*s (124) = .73 and .77, respectively, *p*s < .01, and remains significantly related after controlling for age (Carlson & Harrod, 2013). The scale has successfully distinguished children at risk for poor EF (Fuglestad et al., 2014; Hostinar, Stellern, Schaefer, Carlson, & Gunnar, 2012), and the MEFS in kindergarten predicted first-grade math achievement over and above kindergarten math scores (Hassinger-Das, Jordan, Glutting, Irwin, & Dyson, 2014). In this table-top version, children sorted cards into two opaque boxes affixed with target cards. There were seven levels of complexity, with 3.5-year-olds starting at Level 3 and 4.5-year-olds starting at Level 4. Level 4 was comparable to the standard DCCS and involved sorting red stars and blue trucks into boxes with target cards of a blue star and a red truck, first by color (five trials preswitch), and then by shape (five trials postswitch). Level 3 involved sorting cards where the dimensions of color and shape were separated (e.g., a black star on red background, a black truck on blue background). If children passed the initial level, scoring at least 4 of 5 on both preswitch and postswitch, more difficult levels were administered until they failed. If children failed the initial level, lower levels were administered in a backward fashion until they passed. In this manner, the highest level at which children could pass was determined (Carlson & Schaefer, 2012; Carlson & Zelazo, 2014).

Results

Descriptives

An alpha level of .05 was used for all statistical analyses. No statistical differences were found between the experimental, control-risk, and control-no-risk groups with respect to age, gender, baseline delaying, PPVT, or MEFS, and therefore the groups were combined for baseline analyses. Marginally more 4.5-year-olds were categorized as delayers than 3.5-year-olds, $\chi^2(1, N = 140) = 2.86$, $p = .065$, $\phi = .14$, and more females were categorized as delayers than males, $\chi^2(1, N = 140) = 8.08$, $p = .004$, $\phi = .24$ (see Table 1). No relation was seen with PPVT, $t(136) = -0.04$, $p = .965$.

Switching

In the control-risk group ($n = 33$), only 5 children switched spontaneously (15.2%), and in the control-no-risk group ($n = 30$), only 3 children did so (10.0%). Given that the two groups did not differ in the rate of switching, $\chi^2(1, N = 63) = 0.38$, $p = .410$, or on other variables described above, the two groups were collapsed into a single control group for further analyses ($n = 63$). In the experimental group ($n = 77$), 27 children switched (35.1%), which was significantly greater than the 8 children (12.7%) who spontaneously switched in the control group, $\chi^2(1, N = 140) = 9.24$, $p = .002$, $\phi = .26$.

To explore the nature of children's switching in the experimental group, we examined how children's responses changed across the nine trials. First, switchers and nonswitchers could be differentiated by the first trial of the modified delay task. Delayers who switched were less likely to delay on the first trial ($M = 0.44$, $SD = 0.51$) compared to those who did not switch ($M = 0.88$, $SD = 0.33$), $t(31) = 2.98$ $p = .006$, $d = 1.02$. Similarly, nondelayers who switched were more likely to delay on the first trial ($M = 0.36$, $SD = 0.50$) compared to those who did not switch ($M = 0.03$, $SD = 0.17$), $t(42) = -3.31$, $p = .002$, $d = .88$.

However, all children in the experimental group, including those who did not meet criterion as switchers, showed change between the first and second half of trials (Block 1: Trials 1–4; Block 2: Trials 6–9; Trial 5 omitted). In a repeated measures

Table 1
Baseline Behaviors in Delaying and Saving

	Study 1 (Delaying)					
	3.5 years		4.5 years		Total	
Sex	D	ND	D	ND	D	ND
Male	9 (20.4)	35 (79.5)	11 (33.3)	22 (66.7)	20 (26.0)	57 (74.0)
Female	15 (42.8)	20 (57.1)	16 (57.1)	12 (42.8)	31 (49.2)	32 (50.8)
Total	24 (30.4)	55 (69.6)	27 (44.3)	34 (55.7)	51 (36.4)	89 (63.6)
	Study 2 (saving)					
	3.5 years		4.5 years		Total	
Sex	Sa	Sp	Sa	Sp	Sa	Sp
Male	8 (21.6)	29 (78.4)	14 (42.4)	19 (57.6)	22 (31.4)	48 (68.6)
Female	8 (22.2)	28 (77.8)	7 (19.4)	29 (80.6)	15 (20.8)	57 (79.2)
Total	16 (21.9)	57 (78.1)	21 (30.4)	48 (69.6)	37 (26.1)	105 (73.9)

Note. Percentages within age group and sex are presented in parentheses. D = Delayers, ND = Non-delayers, Sa = Savers, Sp = Spenders.

analysis of variance (ANOVA) with baseline delaying as a between-subjects variable, the main effect of block was not significant, but baseline delaying was significant, $F(1, 75) = 13.69$, $p < .001$, $\eta_p^2 = .15$, as was the interaction between baseline delaying and block, $F(1, 75) = 8.42$, $p = .005$, $\eta_p^2 = .10$. Baseline delayers decided to delay less in the second half ($M = 1.91$, $SD = 1.42$) than in the first half ($M = 2.30$, $SD = 1.31$), $t(32) = 2.42$, $p = .021$, $d = .28$, whereas baseline nondelayers decided to delay more in the second half ($M = 1.27$, $SD = 1.44$) than in the first half ($M = 0.89$, $SD = 1.18$), $t(43) = -1.95$, $p = .058$, $d = .29$ (see Figure 1). This suggests learning and adaptation across trials when the context changed, and can be contrasted with no spontaneous increase or decrease in delaying during the baseline administration of the task from the first block ($M = 1.49$, $SD = 1.55$) to second block ($M = 1.53$, $SD = 1.59$), $t(139) = -0.43$, $p = .669$.

In the experimental group, switching was not related to age, $\chi^2(1, N = 77) = 3.50$, $p = .365$ (see Table 2); gender, $\chi^2(1, N = 77) = 2.09$, $p = .114$; or PPVT, $t(75) = 1.32$, $p = .189$. As predicted, switching was related to baseline behavior, with delayers

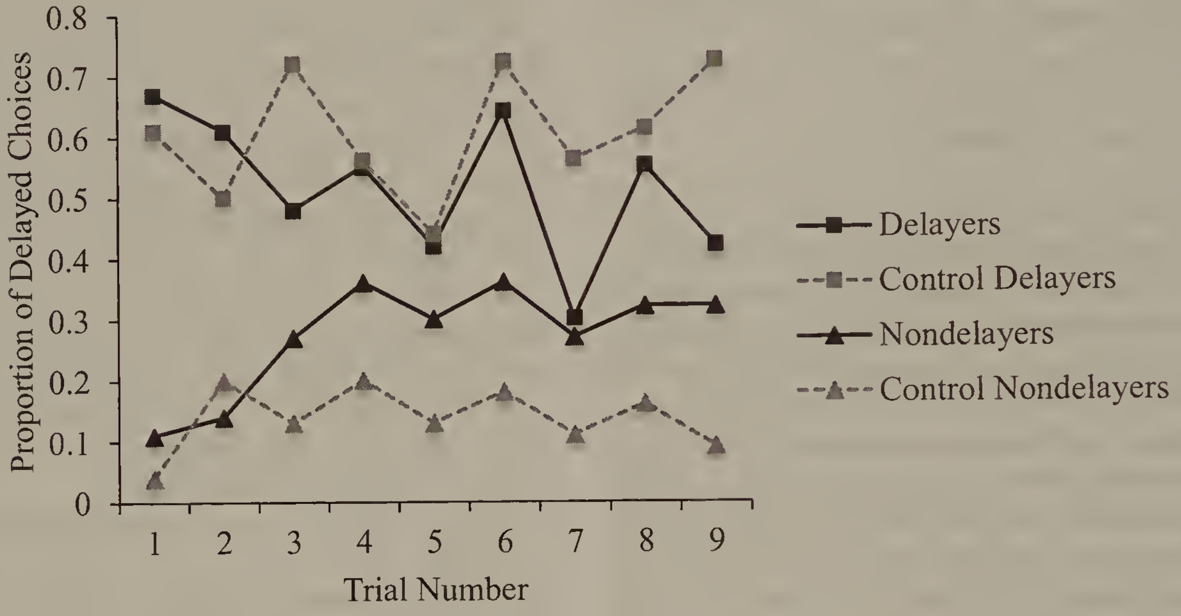


Figure 1. Proportion of delayed choices across trials in the modified delay choice task in Study 1. Separate lines are presented for delayers and nondelayers in the experimental and control conditions.

Table 2
Switching Behavior in Delaying and Saving

	Study 1 (Delaying)					
	Delayers		Non-delayers		Total	
Age	S	NS	S	NS	S	NS
3.5 years	8 (50.0)	8 (50.0)	9 (31.0)	20 (69.0)	17 (37.8)	28 (62.2)
4.5 years	8 (47.1)	9 (52.9)	2 (13.3)	13 (86.7)	10 (31.3)	22 (68.8)
Total	16 (48.5)	17 (51.5)	11 (25.0)	33 (75.0)	27 (35.1)	50 (64.9)
	Study 2 (Saving)					
	Savers		Spenders		Total	
Age	S	NS	S	NS	S	NS
3.5 years	6 (54.5)	5 (45.5)	5 (11.9)	37 (88.1)	11 (20.8)	42 (79.2)
4.5 years	6 (46.2)	7 (53.8)	26 (72.2)	10 (27.8)	32 (65.3)	17 (34.7)
Total	12 (50.0)	12 (50.0)	31 (39.7)	47 (60.3)	43 (42.2)	59 (57.8)

Note. Percentages within age group and baseline behavior are presented in parentheses. S = Switchers, NS = Non-switchers.

(48.5%) switching more than nondelayers (25.0%), $\chi^2(1, N = 77) = 4.57$, $p = .029$, $\phi = .24$ (see Table 2).

To examine whether children recognized which panel was the better choice, children's self-reported preference regarding the two panels was analyzed. Preference data were available for 72 children in the experimental group. Of the 25 children who switched, all but 2 (92.0%) identified the nonrisky panel as their favorite, consistent with their behavioral response. In contrast, of the 47 children who did not switch, 26 (55.3%) favored the risky panel and 21 (44.7%) favored the nonrisky panel, significantly different from the pattern seen with switchers, $\chi^2(1, N = 72) = 15.38$ $p < .001$, $\phi = .46$. Interestingly, almost half of nonswitchers verbally favored the nonrisky panel even though it was inconsistent with their behavior.

MEFS

One of the main hypotheses was that children's switching on the delay task would be positively related to set shifting. An ANOVA was conducted with the highest level passed on the MEFS (0–7) as the dependent variable, and baseline delaying and switching as independent variables, controlling for age and verbal ability, which were significantly related to MEFS performance: age, $F(1, 71) = 39.88$, $p < .001$, $\eta_p^2 = .36$ and PPVT, $F(1, 71) = 9.66$, $p = .003$, $\eta_p^2 = .12$. As hypothesized, baseline delaying was not a significant predictor, $F(1, 71) = 0.84$, $p = .362$; however, switching on the delay choice task was a significant predictor of MEFS performance, $F(1, 71) = 5.09$, $p = .027$, $\eta_p^2 = .07$. Adjusted means controlling for age and PPVT showed that switchers had a mean score of 3.78 ($SE = 0.20$) and nonswitchers had a mean score of 3.21 ($SE = 0.15$). The interaction between baseline delaying and switching was not significant, $F(1, 71) = 1.14$, $p = .290$.

Discussion

We were interested in whether individual differences in preschool children's ability to modify their delay behavior when the context changed would be related to EF. Results showed that some children adapted their behaviors, but this was surprisingly rare. About half of delayers became nondelayers, whereas only one fourth of the nondelayers became delayers, even in the face of losing rewards. This pattern indicates that shifting toward consumption of rewards is easier and requires less control than shifting toward waiting. Furthermore, switchers could be differentiated from nonswitchers by the first trial, indicating that the presentation of new information can be enough to cue a shift in children's behavior. Nonetheless, the latter half of trials showed more rational choices than the first half, suggesting that negative feedback and learning may facilitate switching. Interestingly, even nonswitchers may be learning across trials although they did not meet our criteria for switching. This is consistent with the finding that almost half of nonswitchers reported a preference for the nonrisky choice. It is possible they do not yet reflect sufficiently on the situation to control their behavior (Zelazo & Frye, 1998).

We hypothesized that flexibility of decision making would be associated with age and set-shifting skills. As predicted, children who adjusted their responses on the delay paradigm scored higher on the MEFS, with a small to medium effect size after controlling for age and verbal ability. This suggests that EF skills may be recruited when children must adjust their behavior to the environment in an adaptive manner. Surprisingly, no age differences in switching were observed, although age itself was strongly related to set shifting, and switching was independently related to set shifting. It is possible that a wider age range may reveal developmental differences. Consistent with past findings, baseline delay behavior was not related to verbal ability or set shifting (e.g., Beck et al., 2011; Hongwanishkul et al., 2005; Schwarz et al., 1983).

There are, however, several limitations of the current study. By using risk to manipulate the context, a potential confound was introduced, where nonswitchers may have been attracted to risk taking rather than unable to adapt to the changed context. This concern is mitigated by the lack of switching toward the risky option in the control group. It was also behaviorally difficult to disentangle persistence and extinction (Happaney & Zelazo, 2004). When children fail to switch, they may be perseverating at a no longer rewarded response, or they may be demonstrating persistence in reaching a desired goal. However, children had an alternative response option, and with continued trials and negative feedback, expectations of obtaining treats should be diminished. In Study 2, these issues were addressed by employing a saving paradigm that did not involve risk and uncertainty.

Study 2

Our goal was to examine children's flexibility in two different paradigms examining related decision-making behaviors to strengthen our confidence that relations between behavioral flexibility and EF may indicate a broader psychological process. The marbles saving paradigm (Metcalf & Atance, 2011) was used to measure flexibility in children's saving behavior. The delay choice task was included to explore the relation between baseline saving and delaying, which to our knowledge has not been directly assessed in preschoolers. Therefore, we did not have any strong hypothesis as to the relation between the two behaviors at baseline.

The saving task involved giving children a limited number of marbles for a small marble chute game and recording how many marbles were saved for a later, bigger marble chute game. Children who did not save marbles were unable to play with the big marbles game when it was presented as promised, emphasizing the certainty of future need for marbles. We introduced a modification such that children who saved marbles were also unable to play with the big marbles game, which was revealed to be broken, emphasizing the certainty of no future need for marbles. These experiences should encourage children to adjust their subsequent behavior toward saving or spending their marbles when administered the task a second time.

Method

Participants

A new sample of children was recruited from March 2013 to October 2013 from the same population using the same methods and screening criteria described in Study 1. Three 3.5-year-old boys failed to start or complete the tasks due to fatigue, lack of interest, or noncompliance, and were excluded. The final sample consisted of 142 children, including seventy-three 3.5-year-olds (M = 42.13 months, SD = 0.48, 36 females) and sixty-nine 4.5-year-olds (M = 53.91 months, SD = 0.52, 36 females). Similar to Study 1, the majority of children were White and non-Hispanic (88.0%), with the remaining children identified as White and Hispanic, African American, Asian, or mixed ethnicity. Families were predominantly two-caregiver households (95.1%) and the majority of primary caregivers, mostly mothers, reported attaining a bachelor's degree or higher (88.7%). The median and mode household income was $100,000–$124,999.

Procedure and Measures

Children were tested individually in the same lab environment as in Study 1, and parents completed the same demographics questionnaire. Children were administered the standard delay choice task, followed by the marbles saving task to determine their baseline preferences for saving. They were shown both a small marbles game and a big marbles game. Children were told they would play first with the small game and then later with the big game. However, they were given only three marbles, and must save at least one of the marbles during the small game in order to have marbles to play the big game. Given the reported low to moderate level of saving (Metcalf & Atance, 2011), children who saved at least one marble were categorized as savers, and those who saved no marbles were categorized as spenders. Children were randomly assigned to an experimental or control condition.

In the experimental condition, spenders were given the opportunity to play with the big marbles game, but without marbles, they were unable to play. Savers were given the opportunity to play with the big marbles game, but they discovered that it was broken and could not be played. In the control condition, children did not spend time with the big marbles game, and thus received no intervening feedback between the first and second administrations of the task. A total of 102 children were assigned to the experimental condition, including fifty-three 3.5-year-olds (26 females) and forty-nine 4.5-year-olds (25 females). A total of 40 children were assigned to the control condition, including twenty 3.5-year-olds (10 females) and twenty 4.5-year-olds (11 females).

All children then completed a receptive vocabulary test after which the experimenter discovered three new marbles. Children again played with the small marbles game followed by the big marbles game. Using the same criteria as above, children were again categorized as savers or spenders. Children who changed classifications across the two administrations were considered switchers (e.g., saver to spender), whereas those who did not were considered nonswitchers (e.g., remained a saver). At the end of the second saving task, regardless of performance, children were allowed to play with a third, even bigger marbles game with extra marbles, followed by an independent measure of set shifting.

Marbles saving task. This task was adapted from Metcalf and Atance (2011). Children were introduced to a small marble game and a more appealing big marbles game. The experimenter demonstrated how marbles roll into a box where they cannot be retrieved ("Once you put a marble down the hole, it rolls all the way down and it goes into the box. It stays inside and you cannot use it again"). Children were told that they would first play with the small marbles game for 3 min, followed by the big marbles game for 3 min. They were shown a transparent bag with marbles and told there were only three marbles left. The big marbles game was removed from the room. Before handing children the marbles, children's choices were explicitly stated ("You can use some marbles now for the small game. You can also save some marbles for later for the big game"). Children were given the bag of marbles, and the experimenter started a timer for 3 min and retreated to work in a corner. To discourage interaction, the experimenter wore headphones and minimized conversation, replying as needed with neutral statements (e.g., "Remember I need to work right now"). The small marbles game was removed, and the big marbles game was retrieved. Children were given 3 min with the big marbles game.

For savers, a switch was made such that, unbeknownst to the child, a different big marbles game was retrieved that was identical in outward appearance, but on closer inspection, had putty hidden in the tubes, blocking the marble pathways. The experimenter expressed surprise at discovering that the big game was broken, bringing children's attention to the putty. Children were prompted to try a marble and confirm that it would get stuck. Any remaining marbles were collected to prevent children from entertaining themselves with the marbles.

In both cases, children spent 3 min with the big marbles game, but were unable to play with it, either because they lacked marbles, or because the game was broken. In the control group, the experimenter told children there was a mistake and they had to play another game first before the big marbles game. Children did not spend time with the big marbles game and were directly administered the PPVT. Afterward, the experimenter discovered three more marbles in an otherwise empty drawer and the task was repeated. Regardless of performance, all children were given extra marbles and had five marbles to play with a surprise, extra-big marbles game.

Repeated tasks. The baseline delay choice task, the PPVT, and the MEFS described in Study 1 were repeated in Study 2 following the same procedures.

Results

Descriptives

An alpha level of .05 was used for all statistical analyses. No statistical differences were found between the experimental and control groups with respect to age, gender, baseline delaying, baseline saving, PPVT, or MEFS, and therefore, the groups were combined for baseline analyses. For comparison with baseline delaying in Study 1, children scoring in the midrange were omitted from delay analyses (n = 37) and the remaining children were classified as delayers (6–9 trials) and nondelayers (0–3 trials). Of the remaining 105 children, 47 children (44.8%) were delayers and 58 children (55.2%) were nondelayers. There was a significant age difference, with 27 of 44 children (61.4%) delaying among 4.5-year-olds, and 20 of 61 children (32.8%) delaying among 3.5-year-olds, $\chi^2(1, N = 105) = 8.44$, $p = .003$, $\phi = .28$. Delaying also was marginally related to gender, with 29 of 55 females (52.7%) delaying, compared to only 18 of 50 males (36.0%) delaying, $\chi^2(1, N = 105) = 2.96$, $p = .063$, $\phi = .17$. There was no relation between delay and PPVT scores, $t(102) = -0.46$, $p = .648$.

Baseline saving was unrelated to age, $\chi^2(1, N = 142) = 1.34$, $p = .167$ and gender, $\chi^2(1, N = 142) = 2.07$, $p = .106$ (see Table 1). Baseline saving was also unrelated to baseline delaying, $\chi^2(1, N = 105) = 1.65$, $p = .148$. Savers did, however, score significantly higher on PPVT ($M = 127.28$, $SD = 10.82$) than did spenders ($M = 117.95$, $SD = 13.28$), $t(138) = -3.80$, $p < .001$, $d = .77$.

Switching

In the experimental group, 43 of 102 children (42.2%) switched, which was significantly higher than the 5 of 40 children (12.5%) who spontaneously switched in the control group, $\chi^2(1, N = 142) = 11.29$, $p < .001$, $\phi = .28$. This suggests that the manipulation was effective and in the absence of negative consequences, children's saving and spending behavior was highly stable. In the experimental group, more 4.5-year-olds switched than 3.5-year-olds, $\chi^2(1, N = 102) = 20.72$, $p < .001$, $\phi = .45$ (see Table 2). Marginally more females switched, with 26 of 51 females (51.0%) switching, compared to 17 of 51 males (33.3%), $\chi^2(1, N = 102) = 3.26$, $p = .054$, $\phi = .18$. Although not significant, the proportion of switching was higher when switching from saving to spending than the reverse, $\chi^2(1, N = 102) = 0.79$, $p = .256$ (see Table 2). Switchers scored higher on PPVT ($M = 123.37$, $SD = 13.39$) than nonswitchers ($M = 117.68$, $SD = 12.99$), $t(100) = -2.16$, $p = .033$, $d = .43$.

To examine whether children recognized which was the adaptive choice, children's self-reported preferences for playing the small game or waiting for the big game were analyzed. Preference data were available for 98 children in the experimental group. Of the 12 savers who switched, 9 (75.0%) preferred to play with the small game; of the 11 savers who did not switch, 6 (54.5%) also preferred to play with the small game. Of the 31 spenders who switched, 29 (93.5%) preferred waiting for the big game; of the 44 spenders who did not switch, 36 (81.8%) also preferred waiting for the big game. In total, only 5 (11.6%) of the 43 switchers reported an inconsistent preference, whereas 42 (76.4%) of the 55 nonswitchers reported an inconsistent preference, $\chi^2(1, N = 98) = 40.52$, $p < .001$. These results are similar to Study 1, suggesting that some children recognized the better choice, despite not demonstrating it in behavior.

MEFS

As in Study 1, the main hypothesis was that switching on the saving task would be positively related to an independent measure of set shifting. An ANOVA was conducted with the highest level passed on the MEFS (0–7) as the dependent variable, and baseline saving and switching as independent variables, controlling for age and verbal ability, which were significantly related to MEFS performance: age, $F(1, 96) = 47.05$, $p < .001$, $\eta_p^2 = .33$, and PPVT, $F(1, 96) = 4.24$, $p = .042$, $\eta_p^2 = .04$. Baseline saving was a marginally significant predictor, $F(1, 96) = 3.04$, $p = .084$, $\eta_p^2 = .03$, with adjusted means of 3.98 for savers (*SE* = 0.22) and 3.55 for spenders (*SE* = 0.12). In contrast, controlling for both age and PPVT, switching on the saving task was a significant predictor of MEFS performance, $F(1, 96) = 6.43$, $p = .013$, $\eta_p^2 = .06$. Adjusted means controlling for age and PPVT were 4.09 of 7 (*SE* = 0.19) for switchers and 3.45 (*SE* = 0.17) for nonswitchers. The interaction between switching and baseline saving was marginally significant, $F(1, 96) = 2.98$ $p = .088$, $\eta_p^2 = .03$.

Given the marginally significant interaction, separate ANOVAs, controlling for age and PPVT were conducted for savers and spenders. Switching was not a significant predictor of MEFS performance in baseline savers, $F(1, 20) = 0.19$, $p = .665$, with adjusted means of 4.30 (*SE* = 0.29) for switchers and 4.12 (*SE* = 0.29) for nonswitchers. In contrast, switching was a significant predictor of MEFS performance in baseline spenders, $F(1, 74) = 9.34$, $p = .003$, $\eta_p^2 = .11$, with adjusted means of 3.95 (*SE* = 0.22) for switchers and 2.99 (*SE* = 0.17) for nonswitchers, a full level advantage on the MEFS (see Figure 2).

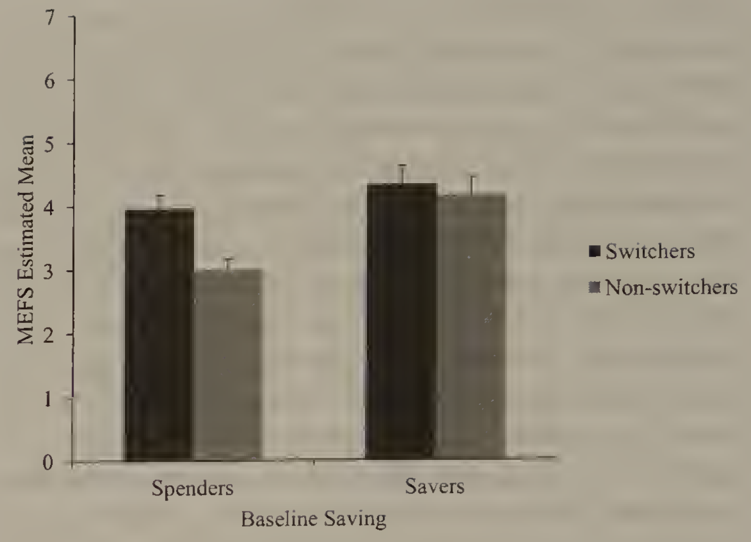


Figure 2. Minnesota Executive Function Scale (MEFS) scores adjusted for age and receptive vocabulary scores in baseline spenders and savers who either switched or did not switch their decisions in Study 2. Bars represent standard errors.

Discussion

We attempted to replicate the findings of Study 1 using a different economic decision-making paradigm. In particular, we were interested in whether children would modify their saving behavior based on past experiences that provided information about the environmental context and whether this would be related to better EF. Results showed that a comparable yet low proportion of children switched their behavior, 35.1% in Study 1 and 42.2% in Study 2. The proportion of spenders who were switchers in Study 2 was 39%, which was the same proportion of spenders who saved during a second trial in Metcalf and Atance (2011).

A greater proportion of children switched from saving to spending than the reverse, which mirrors findings from Study 1, where more children switched from delaying to not delaying. Unlike the delay choice task, which included nine trials, the saving task involved only one opportunity for feedback. Even so, some individuals adjusted their responses, consistent with our finding that switchers and nonswitchers on the modified delay choice task could be differentiated by the first trial. Replicating Study 1, many children who did not switch nevertheless stated a preference for the adaptive choice, although not reflected in their behavior.

Switchers, as before, scored higher on the MEFS, controlling for age and verbal ability, and this relation was stronger in spenders, who needed to shift toward saving. In spenders, the adjusted mean difference between switchers and nonswitchers was a full level on the MEFS, equivalent to 1 year of growth (Carlson, 2012). Unlike Study 1, older children were more likely to switch than younger children. Compared to the delay task, the saving task may require more complex understanding of causal connections between saving and spending and their consequences. Consistent with this view, verbal ability, which may facilitate this understanding, was related to baseline saving and switching. Future-oriented thinking might also play a role, as prior research suggests developmental change during the preschool years (see Atance & O'Neill, 2005). Similar to baseline delaying, baseline saving behavior was also unrelated to set shifting.

Overall, both studies showed parallel patterns of older children and females delaying more often at baseline, a weak relation between baseline behaviors and MEFS performance, greater switching toward immediate rewards, and significant relations between flexibility on a decision-making task and individual differences in set shifting.

General Discussion

The ability to adjust one's behavior flexibly in response to different situations is an important aspect of adaptive functioning. What constitutes "rational" behavior changes with the environment and with current internal goals, states, and motivations. Sometimes it is better to exert control; other times, it is better to let go. As described by Block and Block (1980), individuals may have baseline tendencies toward undercontrol or overcontrol, but the capacity to modulate the level of control in either direction is what comprises ego resiliency, the ability to flexibly respond to changing contingencies.

Across two paradigms investigating decision-making behaviors, preschool children's responses were indeed sensitive to contextual changes and past experience. Furthermore, children were able to flexibly adjust their behavior in both directions, toward delaying and saving, or toward immediate consumption to maximize economic outcomes. More interestingly, however, individual differences emerged, with some children switching to new responses, and others perseverating on old responses. Depending on age and condition, only 25%–50% of children adjusted their decisions, despite receiving consistent and clear negative feedback regarding their original preference. This finding underscores the trenchant stability of delay- and saving-related behavior, which along with immature EF skills, might make young preschoolers especially prone to perseveration. Importantly, the minority of children who did adjust their behavior were differentiated by greater EF, specifically, set shifting, suggesting that EF may underlie adaptation to different environmental demands, whether it requires one to exert or to relinquish control. The degree to which EF is recruited, however, may depend on the difficulty of the behavior required. For example, shifting toward immediate gratification may be less taxing and require less EF as individuals are typically drawn to appetitive rewards and find waiting aversive.

Considering children's well-documented gains in EF during the preschool years (e.g., Carlson, 2005), age was surprisingly not consistently related to switching. Age differences were observed during the saving paradigm, but not during the delay paradigm. Maturation may be most relevant in more complex problems that demand greater understanding of causal connections and contingencies in the situation. Although a consistent relation was observed between EF and switching behavior, this was not the case for EF and baseline behaviors, in keeping with previous findings (e.g., Beck et al., 2011). This pattern of results underscores the primary contribution of this research, which illustrates that being a delayer or a saver is not, in and of itself, indicative of good EF, but rather, learning from experience and adapting one's decisions accordingly is associated with advanced EF skills.

Nature of Children's Switching Behavior

Interestingly, children who adapted their behavior often did so based on one trial of learning; nonetheless, there was evidence of learning over trials on the delay task. Children were more likely to choose adaptively during the second half of the task compared to the first half, presumably from repeated feedback across trials. This pattern was even seen in children who did not meet the criteria for switching. These data suggest that children's adaptive decision-making is likely not a matter of simply switching or not switching, but can be understood in terms of how quickly and how much children switch, how much information and feedback is needed to facilitate behavior change and, as our findings illustrate, their EF skills.

Relatedly, some children who did not show switching behavior seemed to understand which choice was more adaptive. In both studies, over half of children who did not switch behaviorally stated a preference for the adaptive choice. Although this might simply be chance responding, a dissociation between knowledge and using that knowledge in the service of a goal is commonly seen in the DCCS (Zelazo & Frye, 1998). Children know the rules of the new game but are unable to follow through in action. They may have trouble reflecting on the rules and integrating them in a higher order fashion to facilitate switching. Nisan and Koriat (1977) reported similar findings, that over two thirds of children who did not delay answered that a "smart" child would choose to delay, and similarly, Prencipe and Zelazo (2005) observed that 3-year-olds chose immediate smaller rewards for themselves but later larger rewards for another person. Together, these studies suggest a transitional stage in which preschool children know the optimal response but do not yet express it in their decisions. Our findings further add the notion that they may benefit from more trials and feedback.

Limitations and Future Directions

The parallel results of the two studies are encouraging, but several limitations should be considered. First, although our interpretation is that children's set-shifting skills facilitated their switching behavior in the delay and saving tasks, our design does not allow causal conclusions about the role of EF. A step toward clarifying the relation between EF and flexible adaptation of these behaviors in future research would be to demonstrate that training or developmental improvements in EF lead to greater switching in our current paradigms.

Second, we were interested in how children would respond when placed in a situation that demanded a response opposite to their natural preference, and thus it was necessary to categorize children at baseline so they could be exposed to the appropriate conditions. However, our groups were defined by using cutoff criteria that were somewhat arbitrary, and excluded a proportion of children in Study 1. This might also affect the generalizability of the results. Future research should explore measures of delay that can more clearly differentiate delayers and nondelayers, and methods to assess change in a more continuous manner, consistent with our finding that switching is not always all or nothing.

Third, our sample consisted primarily of educated, upper-middle-class, White families, and given the emphasis on adaptation to different contexts, it would be important to examine these effects in children from varying backgrounds, especially when different behaviors may be considered adaptive, and environments may be more or less predictable and changing. A more diverse sample would also increase the generalizability of our findings.

Conclusion

Delay of gratification and saving are complex decision-making behaviors influenced by sociocultural, situational, and individual circumstances. Because not delaying or saving can sometimes be adaptive depending on the context, these behaviors do not always clearly reflect self-control abilities. Our current findings, however, do suggest that children's readiness or ability to successfully adapt their delay and saving behavior according to situational demands and past experience is related to better EF. Executive processes likely play an important role in facilitating flexible behavior, especially when adaptive behaviors or decisions conflict with one's natural tendencies, whether they be hedonic desires or fearful inhibitions.

Development itself has been described as movement toward increasing flexibility, differentiation, and organization, allowing the individual to adapt to the environment with a greater range of means, behavioral patterns, and functions (Sroufe & Rutter, 1984). As children's behavioral repertoire and EF skills develop over time, they confer an increased opportunity to engage with the world and learn from experience in a dynamic, flexible, adaptive, and ultimately successful manner.

References

Ainslie, G. (1975). Specious reward: A behavioral theory of impulsiveness and impulse control. *Psychological Bulletin, 82*, 463–496. doi:10.1037/h0076860

Atance, C. M., & O'Neill, D. K. (2005). The emergence of episodic future thinking in humans. *Learning and Motivation, 36*, 126–144. doi:10.1016/j.lmot.2005.02.003

Ayduk, O., Mendoza-Denton, R., Mischel, W., Downey, G., Peake, P. K., & Rodriguez, M. (2000). Regulating the interpersonal self: Strategic self-regulation for coping with rejection sensitivity. *Journal of Personality and Social Psychology, 79*, 776–792. doi:10.1037/0022-3514.79.5.776

Bandura, A., & Mischel, W. (1965). Modification of self-imposed delay of reward through exposure to live and symbolic models. *Journal of Personality and Social Psychology, 2*, 698–705. doi:10.1037/h0022655

Beck, D. M., Schaefer, C., Pang, K., & Carlson, S. M. (2011). Executive function in preschool children: Test–retest reliability. *Journal of Cognition and Development, 12*, 169–193. doi:10.1080/15248372.2011.563485

Block, J. H., & Block, J. (1980). The role of ego-control and ego-resiliency in the organization of behavior. In W. A. Collins (Ed.), *Development of cognition, affect, and social relations: Minnesota Symposia on Child Psychology* (Vol. 13, pp. 39–101). Hillsdale, NJ: Erlbaum.

Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology, 28,* 595–616. doi:10.1207/s15326942dn2802_3

Carlson, S. M. (2012, February). Scaling the development of executive function in preschool children. In S. M. Carlson (Organizer), *Methodological advances in the study of executive function development.* Paper presented at the themed meeting of the Society for Research in Child Development, Tampa, FL.

Carlson, S. M., & Harrod, J. (2013, April). Validation of the Executive Function Scale for early childhood. In J. Griffin (Chair), *Developing the next generation of preschool outcome measures: The Interagency School Readiness Measurement Consortium.* Poster presented at the biennial meeting of the Society for Research in Child Development, Seattle, WA.

Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and theory of mind. *Child Development, 72,* 1032–1053. doi:10.1111/1467-8624.00333

Carlson, S. M., & Schaefer, C. (2012). *Executive Function Scale for Early Childhood. Test manual.* Minneapolis: University of Minnesota.

Carlson, S. M., White, R. E., & Davis-Unger, A. C. (2014). Evidence for a relation between executive function and pretense representation in preschool children. *Cognitive Development, 29,* 1–16. doi:10.1016/j.cogdev.2013.09.001

Carlson, S. M., & Zelazo, P. D. (2011). The value of control and the influence of values. *Proceedings of the National Academy of Sciences of the United States of America, 108,* 16861–16862. doi:10.1073/pnas.1113235108

Carlson, S. M., & Zelazo, P. D. (2014). *Minnesota Executive Function Scale. Test manual.* St. Paul, MN: Reflection Sciences.

Carlson, S. M., Zelazo, P. D., & Faja, S. (2013). Executive function. In P. D. Zelazo (Ed.), *Oxford handbook of developmental psychology: Body and mind* (Vol. 1, pp. 706–724). New York, NY: Oxford University Press.

Casey, B. J., Somerville, L. H., Gotlib, I. H., Ayduk, O., Franklin, N. T., Askren, M. K., & Shoda, Y. (2011). Behavioral and neural correlates of delay of gratification 40 years later. *Proceedings of the National Academy of Sciences of the United States of America, 108,* 14998–15003. doi:10.1073/pnas.1108561108

Cunningham, W. A., & Zelazo, P. D. (2007). Attitudes and evaluations: A social cognitive neuroscience perspective. *Trends in Cognitive Sciences, 11,* 97–104. doi:10.1016/j.tics.2006.12.005

Dalen, L., Sonuga-Barke, E. J. S., Hall, M., & Remington, B. (2004). Inhibitory deficits, delay aversion and preschool AD/HD: Implications for the dual pathway model. *Neural Plasticity, 11,* 1–11. doi:10.1155/NP.2004.1

Dunn, L. M., & Dunn, D. M. (2007). *PPVT–4 manual.* Bloomington, MN: Pearson.

Fuglestad, A. J., Whitley, M. L., Carlson, S. M., Boys, C. J., Eckerle, J. K., Fink, B. A., & Wozniak, J. R. (2014). Executive functioning deficits in preschool children with fetal alcohol spectrum disorders. *Child Neuropsychology.* Advance online publication. doi:10.1080/09297049.2014.933792

Gallimore, R., Weiss, L. B., & Finney, R. (1974). Cultural differences in delay of gratification: A problem of behavior classification. *Journal of Personality and Social Psychology, 30,* 72–80. doi:10.1037/h0036606

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin, 134,* 31–60. doi:10.1037/0033-2909.134.1.31

Happaney, K., & Zelazo, P. D. (2004). Resistance to extinction: A measure of orbitofrontal function suitable for children? *Brain and Cognition, 55,* 171–184. doi:10.1016/S0278-2626(03)00270-7

Hassinger-Das, B., Jordan, N. C., Glutting, J., Irwin, C., & Dyson, N. (2014). Domain-general mediators of the relation between kindergarten number sense and first-grade mathematics achievement. *Journal of Experimental Child Psychology, 118,* 78–92. doi:10.1016/j.jecp.2013.09.008

Hongwanishkul, D., Happaney, K. R., Lee, W. S. C., & Zelazo, P. D. (2005). Assessment of hot and cool executive function in young children: Age-related changes and individual differences. *Developmental Neuropsychology, 28,* 617–644. doi:10.1207/s15326942dn2802_4

Hostinar, C. E., Stellern, S. S., Schaefer, C. M., Carlson, S. M., & Gunnar, M. R. (2012). The impact of early life adversity on executive function in children adopted internationally from orphanages. *Proceedings of the National Academy of Sciences of the United States of America, 109,* 17208–17212. doi:10.1073/pnas.1121246109

Kidd, C., Palmeri, H., & Aslin, R. N. (2013). Rational snacking: Young children's decision-making on the marshmallow task is moderated by beliefs about environmental reliability. *Cognition, 126,* 109–114. doi:10.1016/j.cognition.2012.08.004

Mahrer, A. R. (1956). The role of expectancy in delayed reinforcement. *Journal of Experimental Psychology, 52,* 101–106. doi:10.1037/h0040837

McGuire, J. T., & Kable, J. W. (2012). Decision makers calibrate behavioral persistence on the basis of time-interval experience. *Cognition, 124,* 216–266. doi:10.1016/j.cognition.2012.03.008

Metcalf, J. L., & Atance, C. M. (2011). Do preschoolers save to benefit their future selves? *Cognitive Development, 26,* 371–382. doi:10.1016/j.cogdev.2011.09.003

Michaelson, L., de la Vega, A., Chatham, C. H., & Munakata, Y. (2013). Delaying gratification depends on social trust. *Frontiers in Psychology, 4,* 355. doi:10.3389/fpsyg.2013.00355

Mischel, W. (1958). Preference for delayed reinforcement: An experimental study of a cultural observation. *Journal of Abnormal and Social Psychology, 56,* 57–61. doi:10.1037/h0041895

Mischel, W. (1966). Theory and research on the antecedents of self-imposed delay of reward. In B. A. Maher (Ed.), *Progress in experimental personality research* (Vol. 3, pp. 85–132). New York, NY: Academic Press.

Mischel, W. (1974). Processes in delay of gratification. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 7, pp. 249–292). San Diego, CA: Academic Press.

Mischel, W., Ayduk, O., Berman, M. G., Casey, B. J., Gotlib, I. H., Jonides, J., & Shoda, Y. (2011). "Willpower" over the life span: Decomposing self-regulation. *Social Cognitive and Affective Neuroscience, 6,* 252–256. doi:10.1093/scan/nsq081

Mischel, W., Shoda, Y., & Peake, P. K. (1988). The nature of adolescent competencies predicted by preschool delay of gratification. *Journal of Personality and Social Psychology, 54,* 687–696. doi:10.1037/0022-3514.54.4.687

Mischel, W., Shoda, Y., & Rodriguez, M. L. (1989). Delay of gratification in children. *Science, 244,* 933–938. doi:10.1126/science.2658056

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., & Howerter, A. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41,* 49–100. doi:10.1006/cogp.1999.0734

Murray, K. T., & Kochanska, G. (2002). Effortful control: Factor structure and relation to externalizing and internalizing behaviors. *Journal of Abnormal Child Psychology, 30,* 503–514. doi:10.1023/A:1019821031523

Nisan, M., & Koriat, A. (1977). Children's actual choices and their conception of the wise choice in a delay-of-gratification situation. *Child Development, 48,* 488–494. doi:10.2307/1128644

Olson, S. L., Schilling, E. M., & Bates, J. E. (1999). Measurement of impulsivity: Construct coherence, longitudinal stability, and relationship with externalizing problems in middle childhood and adolescence. *Journal of Abnormal Child Psychology, 27,* 151–165. doi:10.1023/A:1021915615677

Otto, A. M. C., Schots, P. A. M., Westerman, J. A. J., & Webley, P. (2006). Children's use of saving strategies: An experimental approach. *Journal of Economic Psychology, 27,* 57–72. doi:10.1016/j.joep.2005.06.013

Prencipe, A., & Zelazo, P. D. (2005). Development of affective decision-making for self and other: Evidence for the integration of first- and third-person perspectives. *Psychological Science, 16,* 501–505. doi:10.1111/j.0956-7976.2005.01564.x

Schwarz, J. C., Schrager, J. B., & Lyons, A. E. (1983). Delay of gratification by preschoolers: Evidence for the validity of the choice paradigm. *Child Development, 54,* 620–625. doi:10.2307/1130048

Shoda, Y., Mischel, W., & Peake, P. K. (1990). Predicting adolescent cognitive and self-regulatory competencies from preschool delay of gratification: Identifying diagnostic conditions. *Developmental Psychology, 26,* 978–986. doi:10.1037/0012-1649.26.6.978

Smith-Donald, R., Raver, C. C., Hayes, T., & Richardson, B. (2007). Preliminary construct and concurrent validity of the Preschool Self-Regulation Assessment (PSRA) for field-based research. *Early Childhood Research Quarterly, 22,* 173–187. doi:10.1016/j.ecresq.2007.01.002

Sonuga-Barke, E. J. S., & Webley, P. (1993). *Children's saving: A study in the development of economic behavior.* Hillsdale, NJ: Erlbuam.

Sroufe, L. A., & Rutter, M. (1984). The domain of developmental psychopathology. *Child Development, 55,* 17–29. doi:10.2307/1129832

Strickland, B. R. (1972). Delay of gratification as a function of race of the experimenter. *Journal of Personality and Social Personality, 22,* 108–112. doi:10.1037/h0032385

Thompson, C., Barresi, J., & Moore, C. (1997). The development of future-oriented prudence and altruism in preschoolers. *Cognitive Development, 12,* 199–212. doi:10.1016/S0885-2014(97)90013-7

Wärneryd, K. E. (1999). *The psychology of saving: A study on economic psychology.* Cheltenham, UK: Edward Elgar Publishing.

Willoughby, M., Kupersmidt, J., Voegler-Lee, M., & Bryant, D. (2011). Contributions of hot and cool self-regulation to preschool disruptive behavior and academic achievement. *Developmental Neuropsychology, 36,* 162–180. doi:10.1080/87565641.2010.549980

Zelazo, P. D. (2006). The Dimensional Change Card Sort (DCCS): A method of assessing executive function in children. *Nature Protocols, 1,* 297–301. doi:10.1038/nprot.2006.46

Zelazo, P. D., & Frye, D. (1998). Cognitive complexity and control: II. The development of executive function in childhood. *Current Directions in Psychological Science, 7,* 121–126. doi:10.1111/1467-8721.ep10774761

Zelazo, P. D., Müller, U., Frye, D., & Marcovitch, S. (2003). The development of executive function in early childhood. *Monographs of the Society for Research in Child Development, 68*(3, Serial No. 274).

Child Development, September/October 2015, Volume 86, Number 5, Pages 1449–1457

Hearing Where the Eyes See: Children Use an Irrelevant Visual Cue When Localizing Sounds

Karin Petrini
University of Bath and University College London

Peter R. Jones
University College London

Louise Smith
University of Oxford

Marko Nardini
Durham University and University College London

To reduce sensory uncertainty, humans combine cues from multiple senses. However, in everyday life, many co-occurring cues are irrelevant to the task at hand. How do humans know which cues to ignore? And does this ability change with development? This study shows the ability to ignore cross-modal irrelevant information develops late in childhood. Participants performed a sound discrimination task, with or without an irrelevant visual flash, presented synchronously in front of them. Adults ignored the irrelevant visual information, while 7- to 10-year-olds' responses were biased toward the flash location. The findings show that acquiring mature cue combination mechanisms is a multifaceted process that includes learning to ignore irrelevant cues, as well as to optimally combine relevant cues.

We live in a multisensory world and our perception of it can be greatly improved by integrating multiple sensory cues to reduce sensory uncertainty (Ernst & Banks, 2002; Landy, Maloney, Johnston, & Young, 1995). For example, when crossing a road we can use both the sight and sound of an approaching car to best estimate its position. However, our perception can be deceived if we combine sensory cues that we should not. Magicians have exploited this knowledge for centuries to perform ventriloquism, a stagecraft act in which the voice coming from a person (the ventriloquist) is actually perceived as coming from another source (a puppet moving its mouth together with the voice). Although such erroneous combinations of sensory cues can result in entertaining phenomena, they can also have dangerous consequences —for example, when combining the sound of one approaching car with the sight of another. How do we know which cues to integrate and which to ignore? One way would be to ignore (filter out) sensory cues that are irrelevant to the task at hand. For example, in the example above, we ought to ignore a car parked on the other side of the road when localizing the sound of the approaching car. But how does this selective ability develop, and how does it influence cue combination across the life span?

Based on infant studies we know that soon after birth babies can detect and learn cross-modal spatial correspondences (Bremner et al., 2011). For example, Morrongiello, Fenwick, and Chance (1998) showed that newborn babies expected audiovisual cues to remain collocated in space even when the cues were transposed to a different location. Similarly, many studies have demonstrated that young infants are more sensitive to changes in stimuli that occur concurrently in different sensory modalities than to changes in only one modality (e.g., Bahrick, 1992; Bahrick & Lickliter, 2004; Lewkowicz, 1996, 2000). For example, Bahrick (1992) found that infants as young as 3½ months were more sensitive to temporal changes of an audiovisual stimulus than to temporal changes in an auditory or visual stimulus alone. In other words, as formulated in the intersensory redundancy hypothesis (e.g., Bahrick & Lickliter, 2000, 2002), to learn sensory correspondences young infants appear to rely initially on stimulus properties that are not specific to one modality alone.

This work was supported by the James S. McDonnell Foundation 21st Century Science Scholar in Understanding Human Cognition Program, the NIHR Biomedical Research Centre for Ophthalmology at Moorfields and UCL, and the Nuffield Foundation Undergraduate Research Bursary Scheme. The authors would like also to thank the children and their parents or guardians for their participation.

Correspondence concerning this article should be addressed to Karin Petrini, Department of Psychology, University of Bath, Claverton Down, Bath BA2 7AY, United Kingdom. Electronic mail may be sent to k.petrini@bath.ac.uk.

© 2015 The Authors
Child Development © 2015 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2015/8605-0010
DOI: 10.1111/cdev.12397

Interestingly, however, not all amodal properties seem to have the same relevance for young infants (2–6 months of age), with redundancy in temporal characteristics prevailing over that in spatial characteristics (Morrongiello, Fenwick, & Nutley, 1998). This greater reliance on audiovisual synchronization over spatial colocation is still present in adults, although it is greatly reduced (Lewald & Guski, 2003). Consistent findings in support of the early ability of humans to use redundant multisensory information come from neurophysiological studies with nonhuman animals (Wallace & Stein, 1997; Wallace, Wilkinson, & Stein, 1996), as well as from neurorecording studies with human infants (Hyde, Jones, Porter, & Flom, 2010; Reynolds, Bahrick, Lickliter, & Guy, 2014). For example, studies measuring single-cell responses in kittens (Wallace & Stein, 1997) and event-related potentials in 3-month-old infants (Hyde et al., 2010) have shown increased neural responses for simultaneous audiovisual stimulation compared to either visual or auditory stimulation alone.

Beyond infancy, multisensory processing continues to mature during childhood. For example, recent studies have shown that children are less sensitive to audiovisual asynchrony than adults (Hillock, Powers, & Wallace, 2011; Hillock-Dunn & Wallace, 2012). Furthermore, several independent studies using a Bayesian ideal observer framework have shown that young children, unlike adults, fail to combine multimodal information to improve their precision (Gori, Del Viva, Sandini, & Burr, 2008; Gori, Giuliana, Sandini, & Burr, 2012; Gori, Sandini, & Burr, 2012; Nardini, Bedford, & Mareschal, 2010; Nardini, Begus, & Mareschal, 2012; Nardini, Jones, Bedford, & Braddick, 2008; Petrini, Remark, Smith, & Nardini, 2014). For example, while adults can optimally combine auditory and visual information to improve their precision when performing a spatial discrimination task (Alais & Burr, 2004; Gori, Sandini, et al., 2012), children younger than 12 years of age rely predominantly on visual information (Gori, Sandini, et al., 2012). However, this visual dominance appears to be age and task dependent, since studies examining infants' predictions based on sensory priming (Robinson & Sloutsky, 2004), and *children's* temporal discriminations (Gori, Sandini, et al., 2012), report a converse auditory dominance.

Reconciling the high sensitivity to intersensory redundancy found in infants (Bahrick, 1992; Bahrick & Lickliter, 2004) with the lack of optimal integration in young children (which in certain instances presents itself as complete reliance upon one sense; e.g., Gori et al., 2008; Gori, Sandini, et al., 2012) is difficult due to a lack of studies exploring this issue. However, a first step toward a resolution would be to assume that different aspects of cue combination (intersensory redundancy and optimal integration) emerge and take precedence at different ages. That is, if young children (like infants) still rely strongly on temporal intersensory redundancy, then the presence of an irrelevant, but synchronous, visual event should bias their judgment of sound discrimination. This in turn would result in children exhibiting *suboptimal* integration (due to children attributing too much weight to the irrelevant but synchronous information). Such an account would also fit well with the recalibration theory proposed by Gori and colleagues (e.g., Gori et al., 2008; Gori, Sandini, Martinoli, & Burr, 2010), which states that visual information may be used to calibrate other modalities for spatial tasks (e.g., discriminating objects' orientation) during early childhood. That is, if young children do use visual information to calibrate other sensory cues in spatial tasks, then they would not be able to discount or ignore visual information even when it is irrelevant.

Evidence showing that the ability to filter out irrelevant information increases with age has been limited so far to within-modality studies using flankers or auditory noise (Jones, Moore, & Amitay, 2015; Ridderinkhof, van der Molen, Band, & Bashore, 1997; Shepp & Barrett, 1991; Tipper, Bourque, Anderson, & Brehaut, 1989). Here, we examine how selective the integration of multisensory information is during childhood, by testing whether 7- to 10-year-old children and adults could ignore, and so avoid integrating, irrelevant but synchronized multisensory information. We investigated this using an auditory spatial discrimination task with or without a spatially displaced but temporally synchronous task-irrelevant visual cue. We hypothesized that if children's suboptimal integration does not depend at all on their inability to ignore irrelevant synchronous information, then they should perform well on such a task (i.e., they would attribute little or no weight to the irrelevant information). In contrast, if children's suboptimal integration does in part depend on their inability to ignore the irrelevant synchronous information, then they should perform badly on such a task (i.e., they would *over*-weight the irrelevant information).

Method

Participants

Seventeen children (7–10 years; nine females), and 15 adults (19–38 years; nine females) with normal sight and hearing participated. Participants, recruited by leaflets and press advertisements, were from a socially and ethnically diverse population in central London, UK. The data were collected from June 2012 to September 2013. Sample size was selected based on previous studies using similar paradigms that have fitted psychometric functions to children's data (i.e., using a high number of trials with children; see, e.g.., Barutchu et al., 2010; Gori et al., 2008; Gori, Sandini, et al., 2012; Nardini et al., 2008). Adults and children's parents or guardians gave informed consent to participate, and the study received ethical approval from the research ethics board of University College London.

Procedure

Figure 1 shows the experimental setup. Participants were seated comfortably in front of a computer screen at a distance of 1.7 m. Nine loudspeakers were positioned along the azimuth, at a constant distance from both the participant (1.7 m) and each other (13 cm). On each trial participants fixated on a cross (2.5 × 2.5 cm) at the center of the screen, and were then presented with two sounds (1000 Hz pure tones of 75 dB SPL, lasting 100 ms). The *standard* sound was always played through the speaker positioned in the middle (i.e., the fifth speaker; see Figure 1).

The *comparison* sound was played through any one of the nine speakers and its presentation was counterbalanced across trials. The standard and comparison sounds were separated by 300 ms and their order of presentation was counterbalanced in pseudorandom order across the whole experiment (i.e., in half of the trials the standard was played first and then after 300 ms the comparison was played; in the other half of the trials the opposite occurred). The experiment consisted of two blocks (noiseless and noisy) of 180 trials, for a total of 360 trials. Within each block, trials were uniformly distributed across nine sound locations (positions of the speakers) and two visual conditions (presence and absence of the flash), and these conditions were randomly interleaved. During the noisy block, both standard and comparison sounds were combined additively with a random sample of pink noise. Pink noise is a random noise similar to white noise (e.g., the "static" caused by a detuned radio), but more pleasant to listen to due to the attenuation of high-frequency components. The pink noise was played from all nine speakers and had an intensity equal to 1/10 of either the standard and comparison sound. The standard and comparison sounds were identical except for spatial location, as shown in Figure 1. On each trial, the pink noise onset was 300 ms before the onset of the first sound and ended after the second sound. The noisy auditory condition was included for two reasons: (a) to

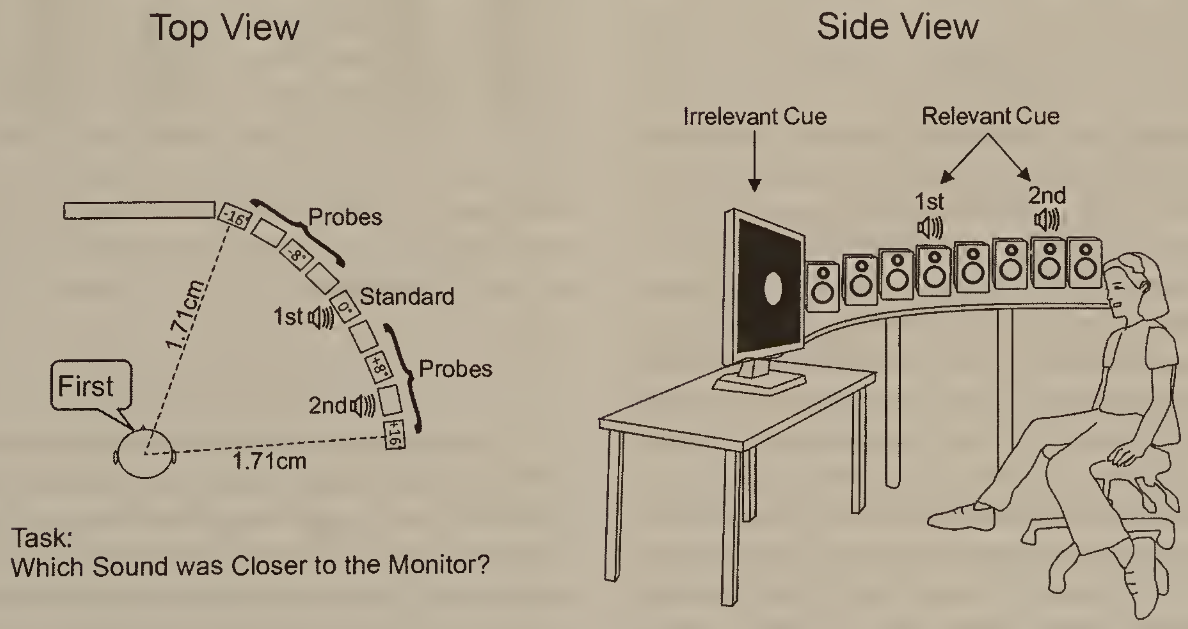


Figure 1. Experimental setup. Nine speakers were positioned at equal distances to the observer, on the right side of a computer screen in a semicircle. The comparison sound could be presented either with or without a synchronous flash (the irrelevant cue), presented directly in front of the observer. The standard sound was always played by the speaker positioned in the middle (see plan view representation on the left). The comparison sound (probe) could be played by any of the nine speakers. The observer judged which sound (the standard or the comparison) was closer to the monitor, either in a noiseless environment or one in which all speakers played added auditory pink noise.

exclude the possibility that age-related differences in the use of the visual irrelevant cue depended solely on differences in sound discrimination threshold and (b) to examine if the use of the visual irrelevant cue increased with the decrease in reliability of the relevant cue (less discriminable sound). The two visual conditions consisted of one in which both the standard and comparison sounds were presented alone, and another in which the standard sound was presented alone while the comparison sound was presented together with a synchronous white flash (17 cm in diameter, 100 ms in duration) in the center of the screen. In each trial, the single flash was perfectly synchronized with the sound in onset, offset, and duration. Synchronization in sound and flash presentation was achieved using a Focusrite Saffire PRO 40 sound card in conjunction with the Matlab Psychtoolbox "PsychPortAudio" ASIO interface (Brainard, 1997; Pelli, 1997).

We asked participants to report verbally which of the two sounds was closer to the monitor. Participants underwent short practice runs separately for the noiseless and noisy conditions. Practice consisted of a few repetitions (a minimum of four, further repeated if needed) of the easiest spatial discrimination trials—those using the first or ninth speaker; Figure 1). This initial practice made sure that the task was clear to participants and that they had experience of the flash as an irrelevant stimulus.

Results

To test whether children, in contrast to adults, would be unable to ignore irrelevant visual information, we asked 7- to 10-year-old children and adults to perform a sound discrimination task, with or without an irrelevant visual flash presented synchronously in front of them. Participants judged whether the first or second tone in a pair was closer to the monitor. In each trial one tone was the standard sound (always played by the speaker positioned in the middle), while the other was the comparison sound (played by any one of the nine speakers). The proportion of trials in which the comparison sound was judged to be closer to the monitor than the standard (0 on the abscissa in Figure 2) was fitted with a cumulative Gaussian independently for each individual. Figure 2a and b (left panels) present the psychometric fits for a representative child and adult participant, obtained using psignifit 2.5.6 (see http://bootstrap-software.org/psignifit/), a software package that implements maximum likelihood parameter estimation. The point at which the psychometric function intersects the 50% correct point on the ordinate is the point of subjective equality. The slope parameter of the functions provides a measure of the spatial discrimination sensitivity (steeper slope = more sensitive). As an exclusion criterion we used the sound-only threshold in both noiseless and noisy conditions (i.e., as that was the parameter of interest, used to determine the relative weight observers gave to auditory and visual information; see Weight Analysis below). The data of two children and two adults were excluded as their performance was at chance level in either sound only (noiseless) or sound only with pink noise (noisy) conditions.

Threshold Analysis

We first examined the effect of age on spatial discrimination thresholds for auditory-only stimuli in the two auditory noise conditions (no noise, pink noise). Figure 2c (left panel) shows that thresholds were higher for children than for adults, and higher with noise than without. A mixed model factorial analysis of variance (ANOVA) with age as a between-subjects factor and noise condition as a within-subjects factor revealed significant main effects of age, $F(1, 30) = 66.217$, $p < .001$, and noise condition, $F(1, 30) = 20.847$, $p < .001$, and a significant interaction, $F(1, 30) = 9.141$, $p = .005$. Paired-samples t tests showed that children had significantly lower thresholds in the noiseless than the noisy condition, $t(16) = -4.404$, $p < .001$, while adults did not differ between conditions, $t(14) = -1.835$, $p = .088$. Independent-samples t tests showed that children had significantly higher thresholds than adults under both noiseless, $t(30) = 4.901$, $p < .001$, and noisy, $t(30) = 8.192$, $p < .001$, conditions.

Weight Analysis

We next examined the effect of age and noise level (without added noise and with added pink noise) on the relative weight attributed to the auditory cue (relevant cue) during the sound + flash condition (see the Supplemental Material for details of how the weights were calculated and for the Matlab code). Figure 2c (right panel) indicates that children weighted the irrelevant visual cue nearly as much as the relevant auditory cue (mean auditory weights just over 0.5). Conversely, adults largely ignored the visual cue (mean auditory weights close to 1). A

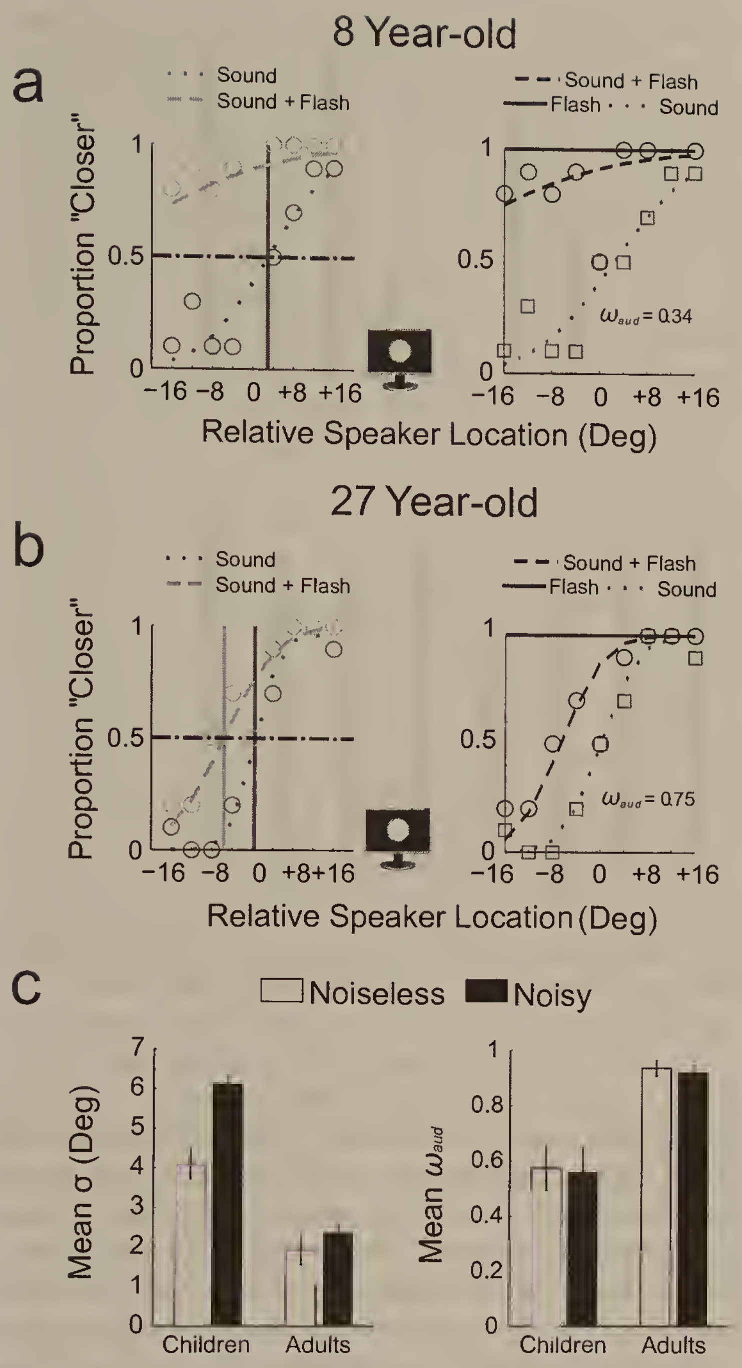


Figure 2. (a). Data from a child in the noiseless condition (1st in Figure 3, top panel). (b). Data from an adult in the noiseless condition (14th adult in Figure 3, bottom panel). Diagrams plot the proportion of "closer" responses given to the comparison sound for two experimental conditions (dashed = sound + flash; dotted = sound only). The monitor image indicates the screen position with respect to the speakers. Solid vertical lines in the left panels represent the point of subjective equality obtained by fitting a cumulative Gaussian function to the data (dashed and dotted lines). For the child, this point fell outside the range tested for the sound + flash condition. The diagrams on the right show the same data, but here the sound + flash is fitted by a model (see equation 1 in the Supplemental Material) estimating the observer's auditory weight ω_{aud}, which determines where this curve lies between those predicted by use of vision only (solid horizontal line on top) versus sound only (dotted line). The child (a; right) is best fitted with an auditory weight of 0.34 (i.e., the dashed sound + flash line is closer to the solid flash line); the adult (b; right) with an auditory weight of 0.75 (i.e., the dashed sound + flash line is closer to the dotted sound line). Since weights must sum to 1, the visual weight for each participant is $1 - \omega_{\text{aud}}$. (c). The diagram on the left plots children's and adults' mean thresholds for the sound only condition (obtained from the slope of the dotted cumulative function) in the auditory noiseless and noisy condition. The diagram on the right plots children's and adults' mean auditory weight (i.e., the weight attributed to the relevant cue) in the auditory noiseless and noisy condition, respectively. The error bars represent the standard error of the mean.

mixed model factorial ANOVA with age as a between-subjects factor and noise condition as within-subjects factor revealed a significant main effect of age, $F(1, 30) = 15.946$, $p < .001$; no main effect of noise condition, $F(1, 30) = 0.372$, $p = .547$; and no interaction, $F(1, 30) < 0.001$, $p = .985$.

The mean weight attributed by children to the auditory cue (the relevant cue) during the sound + flash condition was not significantly different from 0.5 in either the noiseless, $t(16) = 0.917$, $p = .373$, or noisy, $t(16) = 0.619$, $p = .545$, condition. In contrast, the mean weight attributed by adults was significantly different from 0.5 in both noiseless, $t(14) = 15.358$, $p < .001$, and noisy, $t(14) = 15.970$, $p < .001$, conditions. These results indicate that children's greater reliance on irrelevant visual information was not a consequence of their poorer auditory-only threshold. If children's high weighting for vision was explained by a lack of reliability in the auditory cue, then they should have weighted the visual cue even more during the noisy condition. However, although the noisy condition gave higher auditory-only thresholds (Figure 2c, left), children did not weight vision differently in this condition as compared to the noiseless condition (Figure 2c, right).

Correlation Analysis

We used two-tailed Pearson's correlation analyses to ask whether individual children's auditory weights were predicted by their auditory thresholds; they were not, in either noiseless, $r = .044$, $p = .876$, or noisy, $r = -.223$, $p = .424$, conditions. Adults had weightings close to ceiling, one, and so did not have the dispersion needed for a correlation analysis (see Figure 3).

The results indicate that children's weighting of the irrelevant visual cue was not driven by the reliability of the auditory cue. The fact that children do not weight multimodal cues proportionally to their reliability is consistent with previous findings using task-relevant cues (e.g., Gori et al., 2008; Nardini

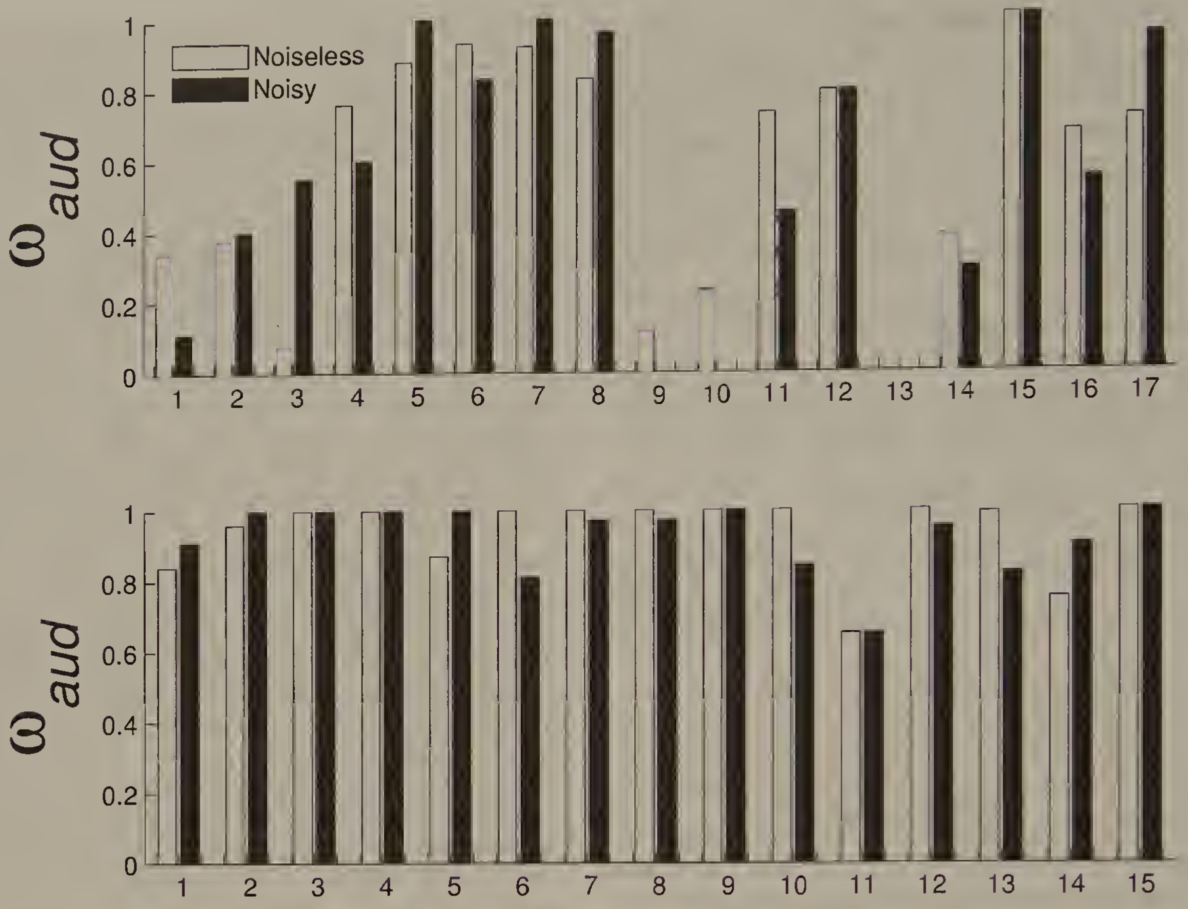


Figure 3. Weight attributed to the auditory cue in the sound + flash condition by each individual child (top panel) and adult (bottom panel) for noiseless (white bars) and noisy (black bars) blocks, respectively. The average data are plotted in Figure 2c right panel.

et al., 2008; Petrini et al., 2014), and with findings demonstrating that decreasing sound reliability by adding noise does not facilitate integration (Barutchu et al. (2010). Finally, a directional (one-tailed) Pearson's correlation analysis revealed that within the child group, the weight attributed to the relevant auditory cue increased with age, $r = .318$, $p = .034$. Bias corrected bootstrap 95% confidence intervals for r were [0.035, 0.580]. As weights sum to 1, this means that the relative weight given to the irrelevant cue decreased with age.

Discussion

The results indicated that children, in contrast to adults, weighted an irrelevant visual cue as much as a relevant auditory cue when discriminating between different locations of a sound. Hence, children performed poorly in a task that required ignoring irrelevant visual information. Furthermore, the weight attributed by children to the irrelevant visual cue decreased with age.

Our findings show that acquiring mature cue combination mechanisms is a multifaceted process and that the late development of optimal cue combination for relevant information (e.g., Gori et al., 2008; Gori, Sandini, et al., 2012; Nardini et al., 2012, 2008; Petrini et al., 2014) is paralleled by a late development of mechanisms selecting cues for combination. Being able to keep sensory information separate (Nardini et al., 2010) may be adaptive when the body is growing, as each sense needs to be continuously recalibrated (Burr, Binda, & Gori, 2011). Our results support the recalibration theory by Gori and colleagues in that keeping cues separate could be adaptive for calibrating other sensory cues with vision, especially in spatial tasks (e.g., Burr et al., 2011; Gori et al., 2008, 2010).

In our task, young children relied on vision in a spatial localization task for which vision was not informative. Their over-weighting of irrelevant visual information demonstrates that children are poor multisensory selectors. Our findings of an immature multisensory selection mechanism in young children are in agreement with the findings of Innes-Brown et al. (2011) in the temporal domain. Innes-Brown et al. found that children were more susceptible than adults to the flash-beep illusion (Shams, Kamitani, & Shimojo, 2000), in that they were more affected by the number of beeps when judging the number of flashes. That is, they relied on auditory information in a visual task for which audition was not informative. In both these studies, children over-relied on the "dominant" modality—the one that would in normal circumstances be the most useful for the task: vision for spatial localization in the present study, and audition for temporal judgments (number of brief events) in Innes-Brown et al.

These results can also be related to those reported for the McGurk effect (i.e., an illusion inducing the perception of a third new utterance "da," by combining the video of a speaker pronouncing "ga" with a superimposed sound of "ba"; McGurk & MacDonald, 1976), for which children have been found to be less susceptible than adults (McGurk & MacDonald, 1976). Children's lower susceptibility to this illusion indicates that they rely relatively more on audition and less on vision. In the McGurk effect, audition is the "dominant" modality in the sense that it would usually tend to be most useful for the task. Overall then, results from all three studies indicate that children's cue selection is less flexible than that of adults, and is consistent with use of a "default" weighting that relies on cues that would normally be expected to be most useful. With development comes the ability to calibrate this reliance on different cues appropriately to specific sensory situations.

During infancy humans combine multiple senses by strongly relying on intersensory redundancy (Bahrick, 1992; Bahrick & Lickliter, 2002, 2004). Infants and later young children find synchronous redundant stimuli across visual and auditory modalities very salient. However, later in childhood multisensory processing continues to be refined in many ways before it becomes adult-like. For example, children become able to weight cues proportionally to their reliability (Gori et al., 2008; Gori, Sandini, et al., 2012 Nardini et al., 2012, 2008; Petrini et al., 2014), and to increase their multisensory selectivity (Hillock et al., 2011; Hillock-Dunn & Wallace, 2012; Innes-Brown et al., 2011) to optimize precision. Here, we show that despite the presence of individual differences among children and adults, which are very common in the developmental literature on cue combination (e.g., Gori et al., 2008; Nardini et al., 2012; Petrini et al., 2014), children's ability to ignore visually irrelevant but temporally redundant information is much poorer. Our findings suggest that during early childhood there is a shift from associating cues based on simple heuristics, such as temporal co-occurrence (mechanism prioritized during infancy; Bahrick & Lickliter, 2000, 2002; Hyde et al., 2010; Lewkowicz, 1996, 2010), to combining cues using more complex integration models, and thus increasing the behavioral gain (Ernst & Banks, 2002) afforded by these previously learned associations. Changes in levels of cue selectivity across the life span could be used as a marker to examine the progression across different stages of cue combination thus filling the currently existing gap in developmental knowledge.

To understand fully how multisensory mechanisms develop during childhood, we must study them in more complex and naturalistic settings where both relevant and irrelevant cues are present. This would also require the inclusion of an additional parameter—determining the extent to which cues are integrated or segregated—to the standard Bayesian model of cues combination (Ernst & Banks, 2002; Landy et al., 1995). The causal Bayesian inference model described by Kording et al. (2007) does take this new parameter into account to create an optimal Bayesian observer that not only infers the best sensory estimates from two sensory signals (i.e., inferring the source location from two sensory signals) but also whether the signals have a common cause. This kind of model may be tested in future to account for the differences between children and adults shown here.

At this stage we do not know whether the filtering of irrelevant sensory cues and the integration of the relevant cues is supported by the same brain mechanisms, and whether they share the same time courses and neural substrates during development. However, the finding that children's suboptimal integration may in part depend on their inability to ignore the irrelevant synchronous information suggests that there may be some overlap in brain processing. Building on recent studies examining the neural correlates of audiovisual speech processing (Reynolds et al., 2014) future neuroimaging and neurophysiological studies could help to answer these fundamental questions, by comparing children's and adults' neural activity in the presence and absence of relevant and irrelevant cross-modal information. Knowing more about sensory selection and integration mechanisms will be of paramount importance for designing successful rehabilitation and treatment for congenital and acquired sensory deficits (Bebko, Weiss, Demark, & Gomez, 2006; Gogate, Maganti, & Perenyi, 2014; Lawson, Ruff, Mccartondaum, Kurtzberg, & Vaughan, 1984). For example, rehabilitation techniques using unimpaired senses to promote calibration during development (Gori, Sandini, Martinoli, & Burr, 2014) could be accompanied by rehabilitation aimed to help the development of multisensory selection mechanisms.

References

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology, 14*, 257–262. doi:10.1016/j.cub.2004.01.029

Bahrick, L. E. (1992). Infants' perceptual differentiation of amodal and modality-specific audio-visual relations. *Journal of Experimental Child Psychology, 53*, 180–199. doi:10.1016/0022-0965(92)90048-B

Bahrick, L. E., & Lickliter, R. (2000). Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Developmental Psychology, 36*, 190–201. doi:10.1037/0012-1649.36.2.190

Bahrick, L. E., & Lickliter, R. (2002). Intersensory redundancy guides early perceptual and cognitive development. In R. Kail (Ed.), *Advances in child development and behavior* (Vol. 30, pp. 153–187). New York, NY: Academic Press.

Bahrick, L. E., & Lickliter, R. (2004). Infants' perception of rhythm and tempo in unimodal and multimodal stimulation: A developmental test of the intersensory redundancy hypothesis. *Cognitive, Affective, & Behavioral Neuroscience, 4*, 137–147.

Barutchu, A., Danaher, J., Crewther, S. G., Innes-Brown, H., Shivdasani, M. N., & Paolini, A. G. (2010). Audiovisual integration in noise by children and adults. *Journal of Experimental Child Psychology, 105*, 38–50. doi:10.1016/j.jecp.2009.08.005

Bebko, J. M., Weiss, J. A., Demark, J. L., & Gomez, P. (2006). Discrimination of temporal synchrony in intermodal events by children with autism and children with developmental disabilities without autism. *Journal of Child Psychology and Psychiatry, 47*, 88–98. doi:10.1111/j.1469-7610.2005.01443.x

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*, 433–436.

Bremner, J. G., Slater, A. M., Johnson, S. P., Mason, U. C., Spring, J., & Bremner, M. E. (2011). Two- to eight-month-old infants' perception of dynamic auditory-visual spatial colocation. *Child Development, 82*, 1210–1223. doi:10.1111/j.1467-8624.2011.01593.x

Burr, D., Binda, P., & Gori, M. (2011). Multisensory integration and calibration in adults and in children. In J. Trommershauser, K. Kording, & M. S. Landy (Eds.), *Sensory cue integration* (pp. 173–194). Oxford, UK: Oxford University Press. doi:10.1093/acprof:oso/9780195387247.003.0010

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415*, 429–433. doi:10.1038/415429a

Gogate, L., Maganti, M., & Perenyi, A. (2014). Preterm and term infants' perception of temporally coordinated syllable-object pairings: Implications for lexical development. *Journal of Speech Language and Hearing Research, 57*, 187–198. doi:10.1044/1092-4388(2013/12-0403)

Gori, M., Del Viva, M., Sandini, G., & Burr, D. C. (2008). Young children do not integrate visual and haptic form information. *Current Biology, 18*, 694–698. doi:10.1016/j.cub.2008.04.036

Gori, M., Giuliana, L., Sandini, G., & Burr, D. (2012). Visual size perception and haptic calibration during development. *Developmental Science, 15*, 854–862. doi:10.1111/j.1467-7687.2012.2012.01183.x

Gori, M., Sandini, G., & Burr, D. (2012). Development of visuo-auditory integration in space and time. *Frontiers in Integrative Neuroscience, 6*, 77. doi:10.3389/fnint.2012.00077

Gori, M., Sandini, G., Martinoli, C., & Burr, D. (2010). Poor haptic orientation discrimination in nonsighted children may reflect disruption of cross-sensory calibration. *Current Biology, 20*, 223–225. doi:10.1016/j.cub. 2009.11.069

Gori, M., Sandini, G., Martinoli, C., & Burr, D. C. (2014). Impairment of auditory spatial localization in congenitally blind human subjects. *Brain, 137*(Pt. 1), 288–293. doi:10.1093/brain/awt311

Hillock, A. R., Powers, A. R., & Wallace, M. T. (2011). Binding of sights and sounds: Age-related changes in multisensory temporal processing. *Neuropsychologia, 49*, 461–467. doi:10.1016/j.neuropsychologia.2010.11.041

Hillock-Dunn, A., & Wallace, M. T. (2012). Developmental changes in the multisensory temporal binding window persist into adolescence. *Developmental Science, 15*, 688–696. doi:10.1111/j.1467-7687.2012.01171.x

Hyde, D. C., Jones, B. L., Porter, C. L., & Flom, R. (2010). Visual stimulation enhances auditory processing in 3-month-old infants and adults. *Developmental Psychobiology, 52*, 181–189. doi:10.1002/dev.20417

Innes-Brown, H., Barutchu, A., Shivdasani, M. N., Crewther, D. P., Grayden, D. B., & Paolini, A. G. (2011). Susceptibility to the flash-beep illusion is increased in children compared to adults. *Developmental Science, 14*, 1089–1099. doi:10.1111/j.1467-7687.2011.01059.x

Jones, P. R., Moore, D. R., & Amitay, S. (2015). Development of auditory selective attention: Why children struggle to hear in noisy environments. *Developmental Psychology, 51*, 353–369. doi:10.1037/a0038570

Kording, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal inference in multisensory perception. *PLoS ONE, 2*, e943. doi:10.1371/journal.pone.0000943

Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision Research, 35*, 389–412. doi:10.1016/0042-6989(94)00176-M

Lawson, K. R., Ruff, H. A., Mccartondaum, C., Kurtzberg, D., & Vaughan, H. G. (1984). Auditory visual responsiveness in full-term and preterm infants. *Developmental Psychology, 20*, 120–127. doi:10.1037//0012-1649.20.1.120

Lewald, J., & Guski, R. (2003). Cross-modal perceptual integration of spatially and temporally disparate auditory and visual stimuli. *Cognitive Brain Research, 16*, 468–478. doi:10.1016/S0926-6410(03)00074-0

Lewkowicz, D. J. (1996). Perception of auditory-visual temporal synchrony in human infants. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 1094–1106. doi:10.1037/0096-1523.22.5.1094

Lewkowicz, D. J. (2000). The development of intersensory temporal perception: An epigenetic systems/limitations view. *Psychological Bulletin, 126*, 281–308. doi:10.1037/0033-2909.126.2.281

Lewkowicz, D. J. (2010). Infant perception of audio-visual speech synchrony. *Developmental Psychology, 46*, 66–77. doi:10.1037/a0015579

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*, 746–748. doi:10.1038/264746a0

Morrongiello, B. A., Fenwick, K. D., & Chance, G. (1998). Crossmodal learning in newborn infants: Inferences about properties of auditory-visual events. *Infant Behavior & Development, 21*, 543–554. doi:10.1016/S0163-6383(98)90028-5

Morrongiello, B. A., Fenwick, K. D., & Nutley, T. (1998). Developmental changes in associations between auditory-visual events. *Infant Behavior & Development, 21*, 613–626. doi:10.1016/S0163-6383(98)90033-9

Nardini, M., Bedford, R., & Mareschal, D. (2010). Fusion of visual cues is not mandatory in children. *Proceedings of the National Academy of Sciences of the United States of America, 107*, 17041–17046. doi:10.1073/pnas.1001699107

Nardini, M., Begus, K., & Mareschal, D. (2012). Multisensory uncertainty reduction for hand localization in children and adults. *Journal of Experimental Psychology: Human Perception and Performance, 39*, 773–787. doi:10.1037/a0030719

Nardini, M., Jones, P., Bedford, R., & Braddick, O. (2008). Development of cue integration in human navigation. *Current Biology, 18*, 689–693. doi:10.1016/j.cub.2008.04.021

Pelli, D. G. (1997). The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*, 437–442. doi:10.1163/156856897X00366

Petrini, K., Remark, A., Smith, L., & Nardini, M. (2014). When vision is not an option: Children's integration of auditory and haptic information is suboptimal. *Developmental Science, 17*, 376–387. doi:10.1111/desc.12127

Reynolds, G. D., Bahrick, L. E., Lickliter, R., & Guy, M. W. (2014). Neural correlates of intersensory processing in 5-month-old infants. *Developmental Psychobiology, 56*, 355–372. doi:10.1002/dev.21104

Ridderinkhof, K. R., van der Molen, M. W., Band, G. P., & Bashore, T. R. (1997). Sources of interference from irrelevant information: A developmental study. *Journal of Experimental Child Psychology, 65*, 315–341. doi:10.1006/jecp.1997.2367

Robinson, C. W., & Sloutsky, V. M. (2004). Auditory dominance and its change in the course of development. *Child Development, 75*, 1387–1401. doi:10.1111/j.1467-8624.2004.00747.x

Shams, L., Kamitani, Y., & Shimojo, S. (2000, December 14). Illusions. What you see is what you hear. *Nature, 408*, 788. doi:10.1038/35048669

Shepp, B. E., & Barrett, S. E. (1991). The development of perceived structure and attention: Evidence from divided and selective attention tasks. *Journal of Experimental Child Psychology, 51*, 434–458. doi:10.1016/0022-0965(91)90087-9

Tipper, S. P., Bourque, T. A., Anderson, S. H., & Brehaut, J. C. (1989). Mechanisms of attention: A developmental study. *Journal of Experimental Child Psychology, 48*, 353–378. doi:10.1016/0022-0965(89)90047-7

Wallace, M. T., & Stein, B. E. (1997). Development of multisensory neurons and multisensory integration in cat superior colliculus. *Journal of Neuroscience, 17*, 2429–2444.

Wallace, M. T., Wilkinson, L. K., & Stein, B. E. (1996). Representation and integration of multiple sensory inputs in primate superior colliculus. *Journal of Neurophysiology, 76*, 1246–1266.

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's website.

Child Development, September/October 2015, Volume 86, Number 5, Pages 1458–1468

Episodic Future Thinking Is Related to Impulsive Decision Making in Healthy Adolescents

Uli Bromberg, Antonius Wiehler, and Jan Peters
University Medical Center Hamburg-Eppendorf

Delay discounting is a stable trait measure of impulsivity. Engaging in episodic future thinking (EFT) can reduce discounting, but whether individual differences in discounting are associated with differences in future thinking remains unclear. Here, this association was tested in healthy adolescents (n = 49, age range = 12–16 years, fluent German speakers, from a large German city). Data collection was between December 2011 and December 2012. Vividness of EFT (assessed via the Autobiographical Memory Interview) was negatively correlated with discounting (r = −.41, 95% CI [−.63, −.13], r^2 = .17). Regression analyses confirmed that this association was stable when controlling for additional variables, including hormonal measures of pubertal maturation and intelligence. EFT may attenuate impulsivity in young people at risk of engaging in problematic behavior.

Adolescence is characterized by an increased engagement in short-sighted behavior, for example, substance abuse, unintentional injury, and unprotected sex (Eaton et al., 2012). Such behavior can be costly and may have long-lasting consequences. For example, the age of alcohol use initiation predicts the degree of alcohol misuse later in life (Grant & Dawson, 1997). Understanding factors that contribute to impulsive decision making during adolescence is therefore crucial.

One central measure of impulsivity is the degree of delay discounting (DD), the phenomenon that humans (and many animals) devalue future rewards as a function of the time to their delivery. In DD paradigms, participants make repeated choices between smaller–sooner (SS) rewards and larger–later (LL) rewards. While people share a common preference to choose larger over smaller rewards, the delay of the larger reward induces the need for a trade-off between reward value and time. Importantly, increased choices of SS over LL rewards (i.e., steeper discounting of value over time) is typically taken to reflect impulsivity: Individuals who suffer from a number of different disorders of impulse control (e.g., substance abuse, pathological gambling, attention deficit/hyperactivity) typically discount future rewards more steeply than controls (Madden, Petry, Badger, & Bickel, 1997; Mitchell, Fields, D'Esposito, & Boettiger, 2005). In this sense, discounting behavior is a reliable correlate of unhealthy or problematic decision making that has been suggested as a behavioral marker for addiction (Bickel, Koffarnus, Moody, & Wilson, 2014; Story, Vlaev, Seymour, Darzi, & Dolan, 2014).

Many psychiatric disorders including substance abuse emerge during adolescence (Paus, Keshavan, & Giedd, 2008). This is consistent with a peak in impulsive behavior during this period, which is also reflected in steeper discounting in adolescents compared to young adults (De Water, Cillessen, & Scheres, 2014; Olson, Hooper, Collins, & Luciana, 2007). At the same time, several desirable outcomes, such as academic success, sense of self-worth, and the ability to cope with stress are positively correlated with the ability to delay gratification (Casey et al., 2011), a process that is conceptually related to DD, but not identical.

Individual discounting behavior is reasonably stable over time (Ohmura, Takahashi, Kitamura, & Wehr, 2006; Peters & Büchel, 2009) and across different elicitation methods (Odum, 2011). For these reasons, discounting is often considered a trait-like characteristic of an individual (Odum, 2011; Peters & Büchel, 2011). At the same time, contextual factors can change discounting preferences relative to an individual's baseline (Koffarnus, Jarmolowicz,

---

This research was supported by Grant PE1627/3-1 to Jan Peters from the Deutsche Forschungsgemeinschaft.

Correspondence concerning this article should be addressed to Uli Bromberg, Department of Systems Neuroscience, University Medical Center Hamburg-Eppendorf, Hamburg, Germany. Electronic mail may be sent to u.bromberg@uke.de.

© 2015 The Authors
Child Development © 2015 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2015/8605-0011
DOI: 10.1111/cdev.12390

Mueller, & Bickel, 2013; Peters & Büchel, 2011). One such factor that can reduce discounting is episodic future thinking (EFT; Benoit, Gilbert, & Burgess, 2011; Daniel, Stanton, & Epstein, 2013; Peters & Büchel, 2010).

EFT, which is the capacity to vividly imagine personal future events, is closely linked to episodic memory (Schacter, Addis, & Buckner, 2007). For instance, when participants are asked to report details about personal past events from their memory as well as details about events that might happen to them in the future, the number of past and future event details are highly correlated (Addis, Wong, & Schacter, 2008). Also, disorders typically associated with impairments in episodic memory (e.g., Alzheimers disease or hippocampal lesions) are also associated with impaired EFT (Addis, Sacchetti, Ally, Budson, & Schacter, 2009). This capacity to create foresight based on memory has been suggested to support the ability of humans to override immediate impulses, supporting advantageous long-term decisions (Bar, 2009). When DD choices are paired with personal episodic information, participants discount less (Daniel et al., 2013; Lin & Epstein, 2014; Peters & Büchel, 2010). Since the development and pursuit of long-term goals is one important ability that needs to be attained, specifically during the highly impulsive life phase of adolescence (Crone & Dahl, 2012), the relation between the capability to create foresight and impulsive behavior in adolescents might be informative regarding potential ways to reduce impulsive behavior during adolescence.

Few studies have investigated EFT capabilities in adolescents. One study examined narratives about autobiographical events in children (Bohn & Berntsen, 2013). Class grade predicted both the length and coherence of future event narratives, with older children producing significantly longer and more coherent narratives. The Autobiographical Memory Interview (AMI) provides a standardized system that allows the categorization of participant's verbal reports on AM into internal and external event details. The number of details provides a score that reflects the ability to construct memory as well as future events (Levine, Svoboda, Hay, Winocur, & Moscovitch, 2002). Accordingly, the AMI is the current gold standard to assess both AM and EFT (Addis et al., 2008; Addis et al., 2009). Using an adapted version of the AMI, Gott and Lah (2013) found that adolescents report significantly more event details than children for both past and future events, suggesting an increase in the capacity to simulate personal future events with age. However, none of these previous studies have attempted to relate individual differences in EFT to individual differences in impulsivity. Previous studies reported a modulation of discounting behavior via EFT (Daniel et al., 2013; Peters & Büchel, 2010), and this effect was correlated with the vividness of future imagery (Peters & Büchel, 2010). We therefore hypothesized that better (i.e., more detailed) episodic imagery of future events in adolescents might be associated with less impulsive decision making (reduced discounting). This association is of particular interest in the age group studied here, as impulsive behavior and associated adverse consequences are particularly pronounced in adolescents (see above).

Discounting behavior has previously been related to numerical age (De Water et al., 2014; Olson et al., 2007; Scheres et al., 2006; Steinberg et al., 2009). However, during adolescence developmental changes in endocrine levels and physical maturation may be more closely related to impulsivity, risk taking, and sensation seeking than numerical age. For example, Martin et al. (2002) showed that both girls and boys who score higher on self- and parent-reported physical maturation have higher ratings of sensation seeking and increased drug use. On the other hand De Water et al. (2014) did not find physical maturation scores to relate to discounting behavior. In adults both sensation seeking and risk taking have been related to gonadal hormonal levels (Daitzman & Zuckerman, 1980; Stanton, Liening, & Schultheiss, 2011). It has been suggested that pubertal increases in reproductive hormones may increase motivational tendencies (Forbes & Dahl, 2010), possibly mediated by functional and structural development of brain networks. In support of this view, neural reward responses are related to testosterone levels (Forbes et al., 2010). Peper, Koolschijn, and Crone (2013) reported a positive association between testosterone level and risk-taking behavior in adolescents and young adults to be mediated by orbital frontal cortex morphology. In the light of these findings, it is of particular importance in developmental studies to control for individual differences in these developmental measures beyond numerical age (Dorn & Biro, 2011; Forbes & Dahl, 2010). To address this issue, we included measures of circulating testosterone and a self-report questionnaire on physical development (PD).

Method

Participants

Forty-nine adolescents (age = 12–16 years) participated in this study. The number of participants was based on previous behavioral findings where the influence of EFT on DD was found in a group of 30 adults (Peters & Büchel, 2010). Aiming to represent the developmental aspect in this adolescent sample, a somewhat larger *n* (49) was chosen for this study. Each age group (12, 13, 14, 15, and 16 years of age) was represented by five boys and five girls, except for the 12-year-old participants who were represented by five girls and four boys (due to missing further available male 12-year-old participants). All participants as well as their parents or legal guardians provided informed written consent, and the study procedure was approved by the local ethics committee. All participants were fluent German speakers, predominantly from Caucasian families, recruited from two schools in middle- to upper-class areas in a large city in northern Germany. Data were collected between December 2011 and December 2012. For recruitment, the study content was presented to seventh through ninth graders while in class. Recruitment packages to take home were offered to all students who showed interest. All families who subsequently signed up by mail were contacted by telephone and were screened for exclusion criteria. These included any diagnosed psychiatric disorders or diagnosed intellect below 70 IQ-score. The physical health of participants was also screened. This, however, did not automatically imply an exclusion.

Autobiographical Memory Interview

EFT and AM were assessed using an adapted version of the AMI (Levine et al., 2002). The same researcher (U.B.) conducted all interviews following a standardized protocol. Each participant was instructed to report 12 personal episodic events (3 events within the next 6 months, 3 events during the following school year, 3 events within the last 6 months, and 3 events during the previous school year, from three different settings: 4 events related to family life, 4 events related to school life, and 4 events related to their spare time activities). Participants were instructed to elaborate on these events in as much detail as possible. The time frames and settings used for the adolescent population were inspired by Piolino et al. (2007). At the beginning of the interview, general instruction, including examples, was given, emphasizing the importance of the events being both specific and personal, as well as the importance of reporting as many details about the events as possible. Following instructions and the examples, each event was then prompted individually (e.g., "Please tell me the details of one event that is related to your family and which you can imagine to take place within the next 6 months"). Following this prompt, the participant was allowed to freely elaborate on the details pertaining to the particular event for 3 min without interruption. If the participant came to a natural end or the 3 min were over, a standardized follow-up question was asked, depending on what had been told already (e.g., "Could you tell me more about where and when the event will take place, who is with you, how you feel and what you think?"; Levine et al., 2002).

Verbal reports were recorded, transcribed, and subsequently rated by an independent rater who was not involved in data collection. The reliability of the rating procedure was tested with Cronbach's alpha by comparing two sets of ratings (comprising all data from two participants) from three different raters, yielding an internal consistency estimate of $\alpha = .94$. The quantitative scores used in the present analysis were based on a segmentation of the transcribed events into different classes of informational entities. The main distinction was between *internal* and *external* details. Internal details were defined as those details that pertained directly to the event described. For example: "I went to school early in the morning." External details were defined as details that did not pertain to the main event in question (external event details), repetitions, and semantic or factual information. An example of a semantic or factual information: "In the mornings I always go to school early." Both internal and external event details were separated into five subcategories: event (event details describe the unfolding of the story, e.g.: "I fell down"), place (any information that involves localization in space, e.g.: "We went to Spain"), time (life epoch, year, season, month, date, day of week, time of day, or clock time, e.g.: "It was in spring"), perception (auditory, olfactory, tactile/pain, taste, visual, spatial–temporal, e.g.: "There were lit candles everywhere"), and emotion or thought (any detail that pertain to the mental state of the subject at the time of the event, e.g.: "I was sad"). We then computed compound scores for internal and external details by averaging over these subcategories. In line with previous studies using this procedure (Addis et al., 2008), the main variables of interest were the mean summed internal detail scores, separately assessed for future and past events, which reflect the average episodic richness of the events. Control variables include summed past and future semantic scores.

Delay Discounting

Delay discounting was assessed using a computer-based adaptive procedure (Peters & Büchel, 2009) where individuals were asked to make choices between €10 available immediately and larger delayed rewards. The experiment was implemented using the Presentation software package (Neurobehavioral Systems Inc., Berkeley, CA). The procedure was adaptive to the choices made by the subject. The amount of the delayed reward was reduced in a stepwise manner after two successive choices of the delayed option and increased in a stepwise manner after two successive choices of the immediate option. The algorithm terminated as soon as the difference between the accepted/rejected amounts reached a delay-specific cutoff value, ranging from €0.5 to €4. Using this adaptive procedure, the amounts offered, as well as the number of trials, differed among participants (range [amount] = €10.50–€227, range [number of trials] = 78–128). The delays included were 1, 2, 7, 14, 30, 90, and 180 days. All rewards were hypothetical. For the estimation of individual discount rates we applied maximum likelihood parameter estimation (Peters, Miedl, & Büchel, 2012), rather than curve fitting on the indifference points, fitting the model to all trials from each participant, using optimization procedures implemented in Matlab (MathWorks, Natick, MA, *fminsearch*). We fitted participants' choices with a hyperbolic function of the form $SV = R/(1 + k * D)$—where SV = subjective value; R = reward; D = delay in days; k = discount rate—and softmax action selection ($P_{chosen} = [\exp(SV_{chosen}/temp)]/[\exp(SV_{chosen}/temp) + \exp(SV_{other}/temp)]$)—where P = probability; SVchosen = subjective value of the chosen option; SVother = subjective value of the other available option; *temp* = temperature—providing estimates of each participant's individual discount rate (k). This free parameter is typically taken as a direct measure of impulsivity, such that higher values reflect more impulsive choice behavior.

To test the goodness of our hyperbolic model fit, we refitted both the above-mentioned hyperbolic and an exponential function of the form $SV = R * \exp(-k * D)$ to the obtained indifference points, using nonlinear regression. A Wilcoxon matched-pairs signed-rank test revealed that the hyperbolic model provided a better fit than the exponential model ($n = 45$, hyperbolic model median adjusted $R^2 = .9266$; exponential model median adjusted $R^2 = .8957$; $Z = -3.65$, $p < .001$).

Developmental Stage

Developmental stage was assessed via three measures. First, we examined circulating testosterone levels. Because circulating testosterone levels exhibit rhythmic fluctuations, 3 to 4 saliva samples were collected spanning the 2-hr time frame of testing, which always occurred somewhere between 3 p.m. and 7 p.m. Saliva samples were then pooled and analyzed by an external company (Innovation Beyond Limits, Hamburg, Germany) with a highly analytical sensitive luminescence assay (1.8 pg/ml). The functional sensitivity of the assay was 5 pg/ml; the intra-assay coefficients of variation were 1.47%, 3.24%, and 3.01%; and the interassay coefficients of variation were 6.96%, 3.76%, and 4.04%. Second, self-report on PD was obtained using a computer-based procedure with instructions and questions based on the Pubertal Development Scale (PDS; Petersen, Crockett, Richards, & Boxer, 1988) as well as the Picture Based Interview About Puberty by Dorn and Susman (as cited in Shirtcliff, Dahl, & Pollak, 2009). The self-assessment was supported by images depicting five stages of PD (Biro & Dorn, 2005). The questions for both genders included growth in height, growth of body and pubic hair, changes in skin, and a self-evaluation whether development was early or late. Questions specific to girls included growth of breasts and onset of menstrual cycle. Questions specific to boys included changes in voice, changes in facial hair, and growth of penis. A relative score was calculated for each participant by dividing his or her summed scores by maximum possible scores. Finally, numerical age was assessed in days.

Additional Neuropsychological Measures

Due to the known association of intelligence with temporal discounting (Shamosh & Gray, 2008), intelligence was assessed using the German version of the Matrix Reasoning and Vocabulary subscales from the Wechsler Intelligence Scale for Children (WISC–IV; Daseking, Petermann, & Petermann, 2007). While some studies have shown an association between discounting rates and self-reported impulsivity (Kirby & Finch, 2010), other studies have shown no such effects (Cyders & Coskunpinar, 2012). Nonetheless, to account for potential associations between self-reported and behavioral impulsivity, a questionnaire-based measure of impulsivity was also included, the Barratt Impulsivity Scale (BIS-11) for Adolescents (Hartmann, Rief, & Hilbert, 2011).

Sample and Outliers

Prior to data analysis, seven outliers were detected (i.e., participants scoring beyond ±3 *SD* on any measure): One participant scored below 3 *SD* on the WISC Matrix Reasoning, one participant scored below 3 *SD* on DD (log-transformed *k*), one participant scored below 3 *SD* on the PD, one participant scored below 3 *SD* on the PD and above 3 *SD* on AM, and three participants scored above 3 *SD* on EFT. Furthermore, four participants for whom only background traces of testosterone levels could be detected were excluded.

However, subjects were only excluded when constituting an outlier in the variables included in the respective analysis. This resulted in slightly differing samples for the different analyses: The comparison of EFT and AM included 45 participants (21 males, 24 females). The comparison of EFT and DD included 45 participants (22 males, 23 females). All remaining analyses included 38 participants (19 males, 19 females). All samples remained balanced with respect to age and gender (see Table 1 for detailed age and gender distribution).

Results

Data Analysis

Data analyses were performed using R, version 3.1.1 (R Core Team, 2014) and the R-packages ppcor, version 1.0; lsr, version 0.3.2; plotrix, version 3.5-5; psych, version 1.4.2.3; Rstudio, version 0.98.987; and Matlab (MathWorks, Natick, MA). Due to the highly skewed distribution of discount rates, testosterone levels (TL), and semantic scores, all analyses were conducted on the log transformed *k* parameters, log-transformed testosterone levels, and log-transformed semantic scores. Following log transformation, all measures were less skewed: Shapiro-Wilk (*W*) normality test, discount rate ($n = 45$): before transformation, $W = .3476$, $p < .001$; after transformation, $W = .9302$, $p < .01$. Discount rate ($n = 38$): before transformation, $W = .2856$, $p < .001$; after transformation, $W = .9304$, $p < .05$. Testosterone level ($n = 38$): before transformation, $W = .8055$, $p < .001$; after transformation, $W = .9786$, $p > .10$. Future semantic score ($n = 38$): before transformation, $W = .8713$, $p < .001$; after transformation, $W = 9519$, $p > .10$. Past semantic score ($n = 38$): before transformation, $W = .8383$, $p < .001$; after transformation, $W = .9823$, $p > .10$.

Table 1
Number and Gender of Participants per Analysis

		Analysis		
Gender	Age	EFT-DD	EFT-AM	Remaining
Male	12	4	4	4
	13	4	4	4
	14	5	5	5
	15	5	4	3
	16	4	4	3
Female	12	5	5	3
	13	4	4	3
	14	5	5	4
	15	4	5	4
	16	5	5	5
		$n = 45$	$n = 45$	$n = 38$

Note. EFT-DD = correlation analysis of episodic future thinking with delay discounting; EFT-AM = correlation analysis of episodic future thinking with autobiographical memory; Remaining = all remaining analyses.

In our analyses we used the mean of summed internal detail scores, averaged separately for EFT and AM events. The external event detail scores were primarily assessed with the purpose of identifying any informational entities that did not pertain to the event of interest. These scores are not relevant to our hypotheses. Also, since the occurrences of external scores are by nature very few, they do not allow for inclusion as a control variable in any of our analyses. To clarify any potential impact of internal detail scores for past events (AM) and semantic scores (future and past), both were included as predictors of discounting behavior in control analyses. Raw scores from the two WISC–IV subtests were transformed to standard scores according to the WISC–IV manual for all further analyses. Descriptive statistics for all variables are listed in Table 2.

For Cohen's *d* (*d*) we used the pooled standard deviation as the normalizing term. For the linear models and the squared partial correlations ($r^2_{(ab.c)}$ = squared partial correlation between *a* and *b* while keeping c constant) we used the *z*-transformed scores of all variables involved.

Correlation Analyses

Summed internal detail scores for future and past events were highly correlated, $r_{(43)} = .72$, 95% CI [.57, 1.0], $r^2 = .52$ (Figure 1a). The association was slightly increased when controlling for age and gender, $r_{(EFT\ AM.age\ +\ gender)} = .74$, 95% CI [.57, .85], $r^2_{(EFT\ AM.age\ +\ gender)} = .55$.

Log-transformed discount rates and EFT (i.e., internal details for future events) were negatively correlated, $r_{(43)} = -.41$, 95% CI [−.63, −.13], $r^2 = .17$

Table 2
Descriptive Statistics of all Variables

Variable (n = 38)	M	SD	Range From	Range To
EFT	20.96	6.97	9.00	39.83
AM	26.64	9.04	12.00	47.83
FSem	7.13	5.61	0.83	20.33
PSem	5.49	4.67	0.33	21.33
TL	20.91	21.70	1.33	86.09
PD	0.70	0.09	0.49	0.83
Vocabulary	13.53	2.54	9.00	18.00
Matrix reasoning	11.92	2.25	8.00	16.00
DEV (z scores)	0	0.92	−1.90	1.44
INT (z scores)	0	0.84	−1.32	1.39

	Mdn	Interquartile range From	Interquartile range To
DD (k)	.036	.022	.051

Note. EFT = episodic future thinking; AM = autobiographical memory; FSem = future semantic detail; PSem = past semantic detail; TL = testosterone level; PD = self-reported physical development; DEV = aggregated developmental score; INT = aggregated intelligence score; DD = discountrate (k).

(Figure 1b). That is, participants with vivid episodic imagery also showed less impulsive decision making. In the following, we report a range of control analyses that confirmed this finding.

First, to exclude the possibility that the effects were driven by a relatively small number of participants with relatively high EFT scores, we reran the analysis using Spearman's nonparametric rank-order correlation, which confirmed the effect, $r_{s(43)} = -.39$, 95% CI [−.61, −.11], $r_s^2 = .15$. When controlling for any effect of gender via partial correlation, the relation was confirmed, $r_{s(EFT\ DD.gender)} = -.44$, 95% CI [−.65 −.17], $r_{s(EFT\ DD.gender)}^2 = .19$. There were no substantial differences between males and females in DD (estimated difference = −.25, 95% CI [−1.10, .60], $d = .18$) nor in EFT (estimated difference = −2.80, 95% CI [−7.20, 1.59], $d = .38$; Table 3).

The developmental measures were all positively correlated: age_TL, $r_{(36)} = .36$, 95% CI [.05, .61], $r^2 = .13$; age_PD, $r_{(36)} = .67$, 95% CI [.45, .82], $r^2 = .46$; and TL_PD, $r_{(36)} = .39$, 95% CI [.08, .63], $r^2 = .15$. Next we investigated gender differences on all developmental measures, which are summarized in Table 3. In particular, girls had higher scores on PD (estimated difference = .72, 95% CI [.10, 1.35], $d = .76$), and boys had higher scores on TL (estimated difference = −.63, 95% CI [−1.26, .01], $d = .65$).

Based on the strong correlation between age and PD, we aggregated these two measures into one measure of developmental stage (DEV) by averaging z-transformed raw scores.

Next we inspected our measures of intelligence. As expected, the two subscales were correlated, $r_{(36)} = .42$, 95% CI [.12, .66], $r^2 = .18$, and we aggregated these by averaging z-transformed raw scores, thereby obtaining compound scores of intelligence for further analyses. Again, we tested for gender differences in our aggregated intelligence measure. Since the result (reported in z scores) was minor, males ($M = .13$, $SD = 0.80$) and females ($M = -.13$,

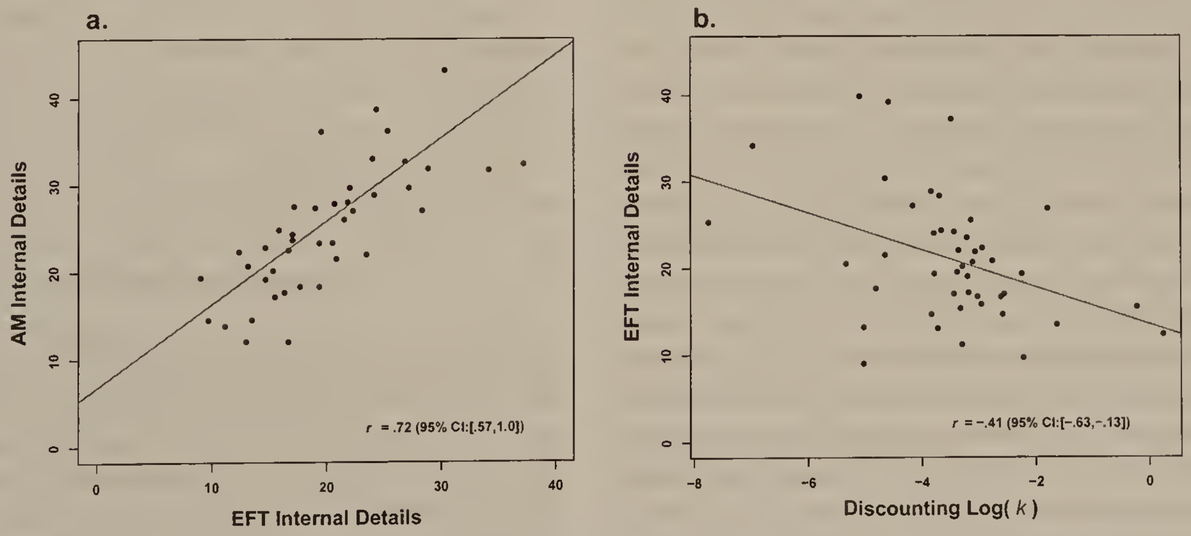


Figure 1. (a) Scatterplots of summed scores for autobiographical memory (AM) versus summed scores for episodic future events (EFT). Higher scores indicate that more episodic details were reported for future and past events, respectively. (b) Discounting behavior as a function of summed scores for episodic future events. Lower log transformed parameter k values indicate less impulsive behavior. Higher episodic future thinking scores indicate that more details were reported for episodic future events.

Table 3
Gender Estimates

	Male		Female		Estimated differences		
Measure	*M*	*SD*	*M*	*SD*	ED	95% CI	*d*
DD	−3.37	1.52	−3.62	1.30	−.25	[−1.10, .60]	.18
EFT	22.45	7.55	19.64	7.04	−2.80	[−7.20, 1.59]	.38
PD	−0.36	1.1	.36	0.72	.72	[.10, 1.35]	.76
TL	0.31	1.2	−.31	0.67	−.63	[−1.26, .01]	.65

Note. ED = estimated difference; *d* = Cohen's *d*; DD = log-transformed discountrate (*k*); EFT = episodic future thinking; TL = z-transformed testosterone level; PD = z-transformed self-reported physical development. For EFT and DD, *n* = 45; for PD and TL, *n* = 38.

SD = 0.88), (estimated difference = −.26, 95% CI [−.82, .30], *d* = .31); gender was not further considered.

Regression Analyses

We then examined the relation between discounting behavior (DD) and EFT while controlling for developmental stage (aggregated numerical age, and PD, DEV), TL, intelligence (aggregated score, INT), questionnaire-based impulsivity (BIS), AM scores, and semantic scores for future as well as past events. To this end, we first set up a reduced linear model including only these control variables as predictors of DD. This model explained approximately 18% of the variability in discount rates (adjusted $R^2 = .18$). Intelligence and TL both showed a trend toward significance (see Table 4, upper part). Next, we extended this to a full model that also included EFT as an additional predictor. The full model explained approximately 30% of the variability in discount rates (adjusted $R^2 = .31$). Effects of EFT and INT were both significant (i.e., the confidence interval did not include zero), while DEV and TL both showed trend-level effects (see Figure 2 and Table 4, lower part). Using a likelihood ratio test we confirmed that the reduction in the residual sum of squares from the reduced to the full model was significant, $\chi^2_{(1,\ n\,=\,38)} = 4.49$, $p < .05$. For both regression models, 11 subjects were excluded due to four missing values in the biological assays and seven outliers in the other predictors. However, it should be noted that only four outliers (8%) were excluded from the initial correlation analysis.

Table 4
Hierarchical Regression

Predictor (*z* scores)	*b**	95% CI	$r^2_{(ab\cdot c)}$
Model predicting DD, excluding EFT (*n* = 38), adjusted $R^2 = .18$			
AM	.013	[−.427, .453]	.00
DEV	−.291	[−.722, .139]	.06
TL	.332	[−.031, .695]	.10
INT	−.433	[−.873, .006]	.12
BIS	.086	[−.276, .448]	.01
FSem	−.172	[−.669, .326]	.02
PSem	−.196	[−.604, .211]	.03
Model predicting DD, including EFT (*n* = 38), adjusted $R^2 = .31$			
AM	.373	[−.125, .871]	.07
DEV	−.363	[−.764, .038]	.11
TL	.307	[−.028, .641]	.11
INT	−.424	[−.829, −.019]	.14
BIS	.060	[−.274, .395]	.00
FSem	−.133	[−.592, .326]	.01
PSem	−.253	[−.631, .125]	.10
EFT	−.450	[−.901, −.098]	.18

Note. Reduced model (upper part); full model (lower part). *b** = standardized beta estimate (point estimate); $r^2_{(ab\cdot c)}$ = squared partial correlation between *a* and *b* while keeping *c* constant (*a* = DD; *b* = the predictor; *c* = all other predictors in the model); DD = log-transformed discountrate (*k*); AM = autobiographical memory; DEV = aggregated developmental score; TL = testosterone level; INT = aggregated intelligence score; BIS = Barratt Impulsiveness Scale; FSem = future semantic detail; PSem = past semantic detail; EFT = episodic future thinking.

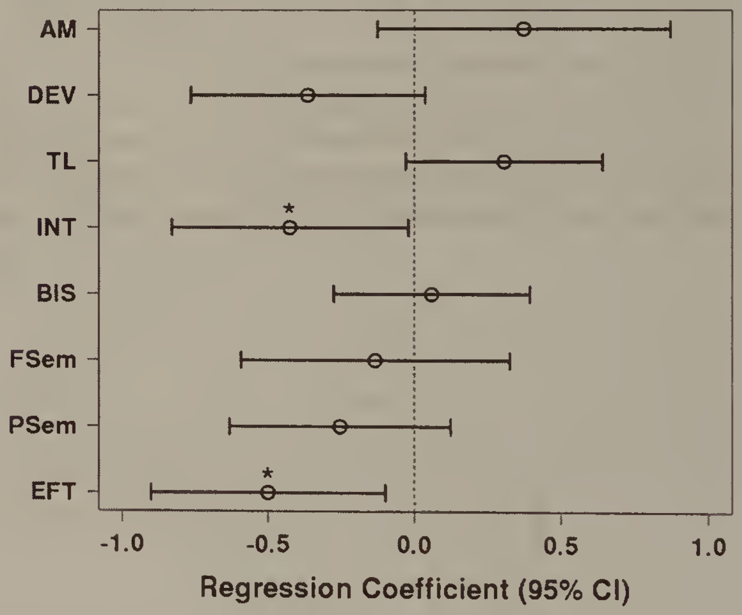


Figure 2. Point estimates and 95% confidence intervals of variables predicting discounting behavior. AM = autobiographical memory; DEV = aggregated developmental scores; TL = testosterone level; INT = aggregated intelligence scores; BIS = Barratt Impulsiveness Scale; FSem = future semantic detail; PSem = past semantic detail; EFT = episodic future thinking.
**p* < .05.

In summary, the relation between discounting behavior and EFT was confirmed even when controlling for a number of potentially relevant variables such as AM, development, intelligence, self-reported BIS impulsivity, and semantic details scores.

Discussion

The present study investigated the association between EFT and DD in adolescents. Besides replicating previous findings (Addis et al., 2008) of a strong positive association between the simulation of past and future events, we most importantly observed a negative association between EFT and DD. Participants who showed more vivid episodic imagery for future events also showed less impulsive (i.e., more future-oriented, patient) choice behavior. Importantly, this relation remained stable even when tested in a regression model controlling for several potential influential factors. Thus, this association appeared to be specifically related to the vividness of *future* imagery of *event* (episodic) details.

Interindividual differences in adolescent's capability to report event details were consistent across temporal directions (i.e., past vs. future events). Neither gender nor age had any effect on this association. This supports the constructive episodic simulation hypothesis (Schacter et al., 2007), which proposes that episodic memories and future simulations are based on a common episodic system, in which memory provides information that can be used to create future simulations. Developmentally, these different aspects of self-projection mature at a similar point in time, namely, during early childhood (Atance & Jackson, 2009). Our data confirm that this relation between past and future events is already established with the onset of adolescence and appears to remain relatively stable during the age range investigated here.

Perhaps the most important finding here is the observation of a negative relation between EFT-scores and discounting behavior in adolescents. EFT explained approximately 17% of the variance in DD. This effect remained stable after controlling for intelligence, physical maturation, testosterone level, AM, semantic details, and self-reported impulsivity, and when the influence of outliers was reduced by applying a nonparametric Spearman rank-order correlation. Previous studies have shown that, in adults, invoking imagery of personal future events during decision making can attenuate discounting (Benoit et al., 2011; Daniel et al., 2013; Palombo, Keane, & Verfaellie, 2014; Peters & Büchel, 2010). In most of these studies, personal future events were associated with future rewards that would be available only at the time of the future events; that is, events and the LL rewards offered were always time-locked. The present findings extend these previous results by showing that baseline EFT capabilities contribute to individual differences in discounting behavior in adolescents. However, since the effect is correlational in nature, our results do not imply a causal relation or specify a direction of influence. One speculative possibility is that high EFT abilities result in a relatively automatic activation of episodic representations during decision making, which then attenuates discounting in a manner that is similar to the previously cited studies that presented episodic information during decision making. Additional research is needed to clarify whether this is the case.

Our results indicate that individual differences in EFT abilities contribute to individual differences in discounting in adolescents, similar to previously reported effects of intelligence and working memory (Bickel, Yi, Landes, Hill, & Baxter, 2011; Shamosh & Gray, 2008; Wesley & Bickel, 2014). In adults, EFT has been related to a range of potential component processes (e.g., working memory and executive functioning; D'Argembeau, Ortoleva, Jumentier, & der Linden, 2010). Accordingly, the association of EFT with discounting behavior might be driven by an executive control component such as working memory (Lin & Epstein, 2014) or intelligence. Similar to previous studies (Shamosh & Gray, 2008), intelligence was a significant predictor of discounting in the present data set. However, our analyses suggest that EFT and intelligence independently explain variance in discounting.

Discounting is a reliable correlate of impulsive behavior and is associated with a wide range of negative health outcomes and problem behaviors, including substance abuse, gambling, and obesity (Bickel et al., 2014; Story et al., 2014). Early onset of such problematic behaviors (e.g., alcohol consumption) is a predictor of later substance use disorders (Grant & Dawson, 1997), and in animal models of deficits in risk-based decision making (Schindler, Tsutsui, & Clark, 2014). EFT ability, on the other hand, is crucial for tasks that attain particular relevance during adolescence. These include creating a life story (Habermas & Bluck, 2000) and developing long-term goals (Crone & Dahl, 2012) and a personal identity (McLean, 2005). Single autobiographical events can be considered the smallest unit of a life story (McLean, 2005) and in this sense EFT abilities might be an important factor supporting these processes during adolescence. In light of the observed association between discounting and EFT, it could be speculated that adolescents may benefit from EFT interventions, similar to adult groups at risk for impulsive health-related decisions (Daniel et al., 2013). However, we acknowledge that our data provide only indirect evidence of a potential association between EFT and problematic behavior in adolescence. EFT predicted DD, which in turn is known to be consistently associated with a range of problematic behaviors (see above). Future studies are needed to test this relation more directly, for example, by focusing on adolescent problem groups such as smokers.

In our model, testosterone levels and a compound score of numerical age and physical maturation showed a trend in predicting discounting behavior. Research on the contribution of endocrine parameters on adolescent impulsive decision making is scarce (Crone & Dahl, 2012; Smith, Chein, & Steinberg, 2013). Two studies in healthy young male students have investigated the relation between discounting behavior and testosterone, reporting inconclusive results. Increased discounting behavior was related to higher testosterone levels only in nonimpulsive subjects (Takahashi, Sakaguchi, Oki, Homma, & Hasegawa, 2006). In contrast, cutaneous administration of a topical testosterone gel did not modulate discounting behavior (Ortner et al., 2013). A recent study showed a positive relation between testosterone level and risk taking (which is somewhat related to impulsivity) in adolescents (Peper et al., 2013). The reported effect was mediated by orbital frontal cortex morphology, which suggests that detection of such relations may require the additional inclusion of neural data.

Despite previous findings of a reduction in discounting with increasing age (Olson et al., 2007), this was evident only as a statistical trend in our data. This might be due to the age range investigated here, which is considerably narrower compared to previous studies that reported age effects in samples including adolescents. Contrary to a previous study that correlated PDS scores with discounting behavior (De Water et al., 2014), we observed a trend-level association between self-reported pubertal maturation and discounting, resembling a previous report of an association between PDS and sensation seeking (Martin et al., 2002). In general, the relations among different measures of pubertal maturation are complex in nature (Shirtcliff et al., 2009), and clarifying any role that differences in pubertal maturation might play in decision making would likely require a larger sample size. However, our analyses show that controlling for differences in pubertal maturation has no impact on the main findings of this study.

Some limitations of the present study need to be acknowledged. Although we accounted for pubertal development beyond age by including testosterone levels as well as the PDS, an examination by a physician (Shirtcliff et al., 2009) was beyond the scope of the present study. Also, the potential influence of the female menstrual cycle on TL was not assessed, which should be taken into consideration when interpreting the endocrine effects. Furthermore, EFT has previously been related to working memory (D'Argembeau et al., 2010). Since we did not assess working memory, we cannot exclude the possibility that the EFT effects might in part reflect differences in working memory. However, fluid intelligence and working memory are typically highly correlated (Diamond, 2013), and intelligence was controlled for in our analyses. Nonetheless, future studies might benefit from a separate assessment of both constructs. Finally, the present study focused on individual differences in adolescents, and future studies might benefit from including additional comparison groups such as younger children and young adults.

Taken together, we show that individual differences in EFT are associated with DD in healthy adolescents. This is of particular interest as this age group is known to be at risk to engage in problematic impulsive behavior (Eaton et al., 2012). Improvement of health-related choice behavior in adolescents, similar to adult groups (Daniel et al., 2013), might help prevent aberrant decision making later in life (Smith et al., 2013; Story et al., 2014).

References

Addis, D. R., Sacchetti, D. C., Ally, B. A., Budson, A. E., & Schacter, D. L. (2009). Episodic simulation of future events is impaired in mild Alzheimer's disease. *Neuropsychologia, 47*, 2660–2671. doi:10.1016/j.neuropsychologia.2009.05.018

Addis, D. R., Wong, A. T., & Schacter, D. L. (2008). Age-related changes in the episodic Simulation of future events. *Psychological Science, 19*, 33–41. doi:10.1111/j.1467-9280.2008.02043.x

Atance, C. M., & Jackson, L. K. (2009). The development and coherence of future-oriented behaviors during the preschool years. *Journal of Experimental Child Psychology, 102*, 379–391. doi:10.1016/j.jecp.2009.01.001

Bar, M. (2009). The proactive brain: Memory for predictions. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*, 1235–1243. doi:10.1098/rstb.2008.0310

Benoit, R. G., Gilbert, S. J., & Burgess, P. W. (2011). A neural mechanism mediating the impact of episodic prospection on farsighted decisions. *Journal of Neuroscience, 31*, 6771–6779. doi:10.1523/JNEUROSCI.6559-10.2011

Bickel, W. K., Koffarnus, M. N., Moody, L., & Wilson, A. G. (2014). The behavioral- and neuro-economic process of temporal discounting: A candidate behavioral marker of addiction. *Neuropharmacology, 76*(Pt B), 518–527. doi:10.1016/j.neuropharm.2013.06.013

Bickel, W. K., Yi, R., Landes, R. D., Hill, P. F., & Baxter, C. (2011). Remember the future: Working memory training decreases delay discounting among stimulant addicts. *Biological Psychiatry, 69*, 260–265. doi:10.1016/j.biopsych.2010.08.017

Biro, F., & Dorn, L. (2005). Puberty and adolescent sexuality. *Pediatric Annals, 34*, 777–784. doi:10.3928/0090-4481-20051001-09

Bohn, A., & Berntsen, D. (2013). The future is bright and predictable: The development of prospective life stories across childhood and adolescence. *Developmental Psychology, 49*, 1232–1241. doi:10.1037/a0030212

Casey, B. J., Somerville, L. H., Gotlib, I. H., Ayduk, O., Franklin, N. T., Askren, M. K., . . . Shoda, Y. (2011). Behavioral and neural correlates of delay of gratification 40 years later. *Proceedings of the National Academy of Sciences of the United States of America, 108*, 14998–15003. doi:10.1073/pnas.1108561108

Crone, E. A., & Dahl, R. E. (2012). Understanding adolescence as a period of social–affective engagement and goal flexibility. *Nature Reviews Neuroscience, 13*, 636–650. doi:10.1038/nrn3313

Cyders, M. A., & Coskunpinar, A. (2012). The relationship between self-report and lab task conceptualizations of impulsivity. *Journal of Research in Personality, 46*, 121–124. doi:10.1016/j.jrp.2011.11.005

Daitzman, R., & Zuckerman, M. (1980). Disinhibitory sensation seeking, personality and gonadal hormones. *Personality and Individual Differences, 1*, 103–110. doi:10.1016/0191-8869(80),90027-6

Daniel, T. O., Stanton, C. M., & Epstein, L. H. (2013). The future is now reducing impulsivity and energy intake using episodic future thinking. *Psychological Science, 24*, 2339–2342. doi:10.1177/0956797613488780

D'Argembeau, A., Ortoleva, C., Jumentier, S., & der Linden, M. V. (2010). Component processes underlying future thinking. *Memory & Cognition, 38*, 809–819. doi:10.3758/MC.38.6.809

Daseking, M., Petermann, U., & Petermann, F. (2007). Intelligenzdiagnostik mit dem HAWIK–IV. *Kindheit Und Entwicklung, 16*, 250–259. doi:10.1026/0942-5403.16.4.250

De Water, E., Cillessen, A. H. N., & Scheres, A. (2014). Distinct age-related differences in temporal discounting and risk taking in adolescents and young adults. *Child Development, 85*, 1881–1897. doi:10.1111/cdev.12245

Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64*, 135–168. doi:10.1146/annurev-psych-113011-143750

Dorn, L. D., & Biro, F. M. (2011). Puberty and its measurement: A decade in review. *Journal of Research on Adolescence, 21*, 180–195. doi:10.1111/j.1532-7795.2010.00722.x

Eaton, D., Kann, L., Kinchen, S., Shanklin, S., Flint, K., Hawkins, J., et al. (2012). Youth risk behavior surveillance—United States, 2011. *Morbidity and Mortality Weekly Report (Surveillance Summaries)*, Washington, DC.

Forbes, E. E., & Dahl, R. E. (2010). Pubertal development and behavior: Hormonal activation of social and motivational tendencies. *Brain and Cognition, 72*, 66–72. doi:10.1016/j.bandc.2009.10.007

Forbes, E. E., Ryan, N. D., Phillips, M. L., Manuck, S. B., Worthman, C. M., Moyles, D. L., . . . Dahl, R. E. (2010). Healthy adolescents' neural response to reward: Associations with puberty, positive affect, and depressive symptoms. *Journal of the American Academy of Child & Adolescent Psychiatry, 49*, 162–172, e5. doi:10.1016/j.jaac.2009.11.006

Gott, C., & Lah, S. (2013). Episodic future thinking in children compared to adolescents. *Child Neuropsychology, 20*, 625–640. doi:10.1080/09297049.2013.840362

Grant, B. F., & Dawson, D. A. (1997). Age at onset of alcohol use and its association with DSM–IV alcohol abuse and dependence: Results from the National Longitudinal Alcohol Epidemiologic Survey. *Journal of Substance Abuse, 9*, 103–110. doi:10.1016/S0899-3289(97),90009-2

Habermas, T., & Bluck, S. (2000). Getting a life: The emergence of the life story in adolescence. *Psychological Bulletin, 126*, 748–769. doi:10.1037/0033-2909.126.5.748

Hartmann, A. S., Rief, W., & Hilbert, A. (2011). Psychometric properties of the German version of the Barratt Impulsiveness Scale, Version 11 (BIS-11) for adolescents. *Perceptual and Motor Skills, 112*, 353–368. doi:10.2466/08.09.10.PMS.112.2.353-368

Kirby, K. N., & Finch, J. C. (2010). The hierarchical structure of self-reported impulsivity. *Personality and Individual Differences, 48*, 704–713. doi:10.1016/j.paid.2010.01.019

Koffarnus, M. N., Jarmolowicz, D. P., Mueller, E. T., & Bickel, W. K. (2013). Changing delay discounting in the light of the competing neurobehavioral decision systems theory: A review. *Journal of the Experimental Analysis of Behavior, 99*, 32–57. doi:10.1002/jeab.2

Levine, B., Svoboda, E., Hay, J. F., Winocur, G., & Moscovitch, M. (2002). Aging and autobiographical memory: Dissociating episodic from semantic retrieval. *Psychology and Aging, 17*, 677–689. doi:10.1037/0882-7974.17.4.677

Lin, H., & Epstein, L. H. (2014). Living in the moment: Effects of time perspective and emotional valence of episodic thinking on delay discounting. *Behavioral Neuroscience, 128*, 12–19. doi:10.1037/a0035705

Madden, G. J., Petry, N. M., Badger, G. J., & Bickel, W. K. (1997). Impulsive and self-control choices in opioid-dependent patients and non-drug-using control patients: Drug and monetary rewards. *Experimental and Clinical Psychopharmacology, 5*, 256–262. doi:10.1037/1064-1297.5.3.256

Martin, C. A., Kelly, T. H., Rayens, M. K., Brogli, B. R., Brenzel, A., Smith, W. J., & Omar, H. A. (2002). Sensation seeking, puberty, and nicotine, alcohol, and marijuana use in adolescence. *Journal of the American Academy of Child & Adolescent Psychiatry, 41*, 1495–1502. doi:10.1097/00004583-200212000-00022

McLean, K. C. (2005). Late adolescent identity development: Narrative meaning making and memory telling.

Developmental Psychology, 41, 683–691. doi:10.1037/0012-1649.41.4.683

Mitchell, J. M., Fields, H. L., D'Esposito, M., & Boettiger, C. A. (2005). Impulsive responding in alcoholics. *Alcoholism: Clinical and Experimental Research, 29*, 2158–2169. doi:10.1097/01.alc.0000191755.63639.4a

Odum, A. L. (2011). Delay discounting: Trait variable? *Behavioural Processes, 87*, 1–9. doi:10.1016/j.beproc.2011.02.007

Ohmura, Y., Takahashi, T., Kitamura, N., & Wehr, P. (2006). Three-month stability of delay and probability discounting measures. *Experimental and Clinical Psychopharmacology, 14*, 318–328. doi:10.1037/1064-1297.14.3.318

Olson, E. A., Hooper, C. J., Collins, P., & Luciana, M. (2007). Adolescents' performance on delay and probability discounting tasks: Contributions of age, intelligence, executive functioning, and self-reported externalizing behavior. *Personality and Individual Differences, 43*, 1886–1897. doi:10.1016/j.paid.2007.06.016

Ortner, G. R., Wibral, M., Becker, A., Dohmen, T., Klingmüller, D., Falk, A., & Weber, B. (2013). No evidence for an effect of testosterone administration on delay discounting in male university students. *Psychoneuroendocrinology, 38*, 1814–1818. doi:10.1016/j.psyneuen.2012.12.014

Palombo, D. J., Keane, M. M., & Verfaellie, M. (2014). The medial temporal lobes are critical for reward-based decision making under conditions that promote episodic future thinking. *Hippocampus, 25*, 345–353. doi:10.1002/hipo.22376

Paus, T., Keshavan, M., & Giedd, J. N. (2008). Why do many psychiatric disorders emerge during adolescence? *Nature Reviews Neuroscience, 9*, 947–957. doi:10.1038/nrn2513

Peper, J. S., Koolschijn, P. C. M. P., & Crone, E. A. (2013). Development of risk taking: Contributions from adolescent testosterone and the orbito-frontal cortex. *Journal of Cognitive Neuroscience, 25*, 2141–2150. doi:10.1162/jocn_a_00445

Peters, J., & Büchel, C. (2009). Overlapping and distinct neural systems code for subjective value during intertemporal and risky decision making. *Journal of Neuroscience, 29*, 15727–15734. doi:10.1523/JNEUROSCI.3489-09.2009

Peters, J., & Büchel, C. (2010). Episodic future thinking reduces reward delay discounting through an enhancement of prefrontal-mediotemporal interactions. *Neuron, 66*, 138–148. doi:10.1016/j.neuron.2010.03.026

Peters, J., & Büchel, C. (2011). The neural mechanisms of inter-temporal decision-making: Understanding variability. *Trends in Cognitive Sciences, 15*, 227–239. doi:10.1016/j.tics.2011.03.002

Peters, J., Miedl, S. F., & Büchel, C. (2012). Formal comparison of dual-parameter temporal discounting models in controls and pathological gamblers. *PLoS ONE, 7*, e47225. doi:10.1371/journal.pone.0047225

Petersen, A. C., Crockett, L., Richards, M., & Boxer, A. (1988). A self-report measure of pubertal status: Reliability, validity, and initial norms. *Journal of Youth and Adolescence, 17*, 117–133. doi:10.1007/BF01537962

Piolino, P., Hisland, M., Ruffeveille, I., Matuszewski, V., Jambaqué, I., & Eustache, F. (2007). Do school-age children remember or know the personal past? *Consciousness and Cognition, 16*, 84–101. doi:10.1016/j.concog.2005.09.010

R Core Team (2014). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from http://www.R-project.org/

Schacter, D. L., Addis, D. R., & Buckner, R. L. (2007). Remembering the past to imagine the future: The prospective brain. *Nature Reviews Neuroscience, 8*, 657–661. doi:10.1038/nrn2213

Scheres, A., Dijkstra, M., Ainslie, E., Balkan, J., Reynolds, B., Sonuga-Barke, E., & Castellanos, F. X. (2006). Temporal and probabilistic discounting of rewards in children and adolescents: Effects of age and ADHD symptoms. *Neuropsychologia, 44*, 2092–2103. doi:10.1016/j.neuropsychologia.2005.10.012

Schindler, A. G., Tsutsui, K. T., & Clark, J. J. (2014). Chronic alcohol intake during adolescence, but not adulthood, promotes persistent deficits in risk-based decision making. *Alcoholism, Clinical and Experimental Research, 38*, 1622–1629. doi:10.1111/acer.12404

Shamosh, N. A., & Gray, J. R. (2008). Delay discounting and intelligence: A meta-analysis. *Intelligence, 36*, 289–305. doi:10.1016/j.intell.2007.09.004

Shirtcliff, E. A., Dahl, R. E., & Pollak, S. D. (2009). Pubertal development: Correspondence between hormonal and physical development. *Child Development, 80*, 327–337. doi:10.1111/j.1467-8624.2009.01263.x

Smith, A. R., Chein, J., & Steinberg, L. (2013). Impact of socio-emotional context, brain development, and pubertal maturation on adolescent risk-taking. *Hormones and Behavior, 64*, 323–332. doi:10.1016/j.yhbeh.2013.03.006

Stanton, S. J., Liening, S. H., & Schultheiss, O. C. (2011). Testosterone is positively associated with risk taking in the Iowa Gambling Task. *Hormones and Behavior, 59*, 252–256. doi:10.1016/j.yhbeh.2010.12.003

Steinberg, L., Graham, S., O'Brien, L., Woolard, J., Cauffman, E., & Banich, M. (2009). Age differences in future orientation and delay discounting. *Child Development, 80*, 28–44. doi:10.1111/j.1467-8624.2008.01244.x

Story, G. W., Vlaev, I., Seymour, B., Darzi, A., & Dolan, R. J. (2014). Does temporal discounting explain unhealthy behavior? A systematic review and reinforcement learning perspective. *Frontiers in Behavioral Neuroscience, 8*, 76. doi:10.3389/fnbeh.2014.00076

Takahashi, T., Sakaguchi, K., Oki, M., Homma, S., & Hasegawa, T. (2006). Testosterone levels and discounting delayed monetary gains and losses in male humans. *Neuro Endocrinology Letters, 27*, 439–444.

Wesley, M. J., & Bickel, W. K. (2014). Remember the future II: Meta-analyses and functional overlap of working memory and delay discounting. *Biological Psychiatry, 75*, 435–448. doi:10.1016/j.biopsych.2013.08.008

Child Development, September/October 2015, Volume 86, Number 5, Pages 1469–1488

Early Nonparental Care and Social Behavior in Elementary School: Support for a Social Group Adaptation Hypothesis

Jean-Baptiste Pingault
University College London

Richard E. Tremblay
University College Dublin and University of Montreal

Frank Vitaro
University of Montreal

Christa Japel
Université du Québec à Montréal

Michel Boivin, and Sylvana M. Côté
Université Laval and Tomsk State University

This study examined the contribution of nonparental child-care services received during the preschool years to the development of social behavior between kindergarten and the end of elementary school with a birth cohort from Québec, Canada (N = 1,544). Mothers reported on the use of child-care services, while elementary school teachers rated children's shyness, social withdrawal, prosociality, opposition, and aggression. Children who received nonparental child-care services were less shy, less socially withdrawn, more oppositional, and more aggressive at school entry (age 6 years). However, these differences disappeared during elementary school as children who received exclusive parental care caught up with those who received nonparental care services. This "catch-up" effect from the perspective of children's adaptation to the social group is discussed.

The use of nonparental child-care services during the preschool years is now the norm rather than the exception and this new reality is raising questions regarding the effect of those services on children's social development (Lamb & Ahnert, 2006; Linting & van IJzendoorn, 2009; Sagi, Koren-Karie, Gini, Ziv, & Joels, 2002). The human and nonhuman attachment literature has emphasized that maternal proximity and availability during infancy are essential for offspring adjustment, suggesting potential adverse effects of child-care services, in particular, intensive child-care services in the 1st year of life (Aviezer & Sagi-Schwartz, 2008; Belsky, 2006; Dallaire & Weinraub, 2007; National Institute of Child Health and Human Development [NICHD] Early Child Care Research Network, 2001; Sagi et al., 2002). However, initial research on attachment has also highlighted the unique importance of peers for social development throughout early childhood (Harlow, 1972). Later research confirmed the crucial role of peers in children's social development, be it beneficial or not (Boivin, Vitaro, & Poulin, 2005; Goldstein, Arnold, Rosenberg, Stowe, & Ortiz, 2001; Rubin, Coplan, & Bowker, 2009). Some authors argued that the social environment in child-care services, and peers' influence in particular, could be one of the mechanisms explaining the effects of child-care services on social behavior (Averdijk, Besemer, Eisner, Bijleveld, & Ribeaud, 2011; Boivin, Vitaro, et al., 2005; Fabes, Hanish, & Martin, 2003; McCartney et al., 2010). We use the term *nonparental child-care services* or *child-care services* to refer to preschool care and education services provided on a regular basis by a person other than the parent.

This study was funded by the Quebec Government, the Canadian Institutes of Health Research, the Social Sciences and Humanities Research Council of Canada, the Quebec Fund for Research on Society and Culture, the Health Research Fund of Quebec, the Lucie and André Chagnon Foundation, the University of Montreal and Laval University. Dr Pingault is supported by a Marie Curie Intra-European Fellowship (330699). We thank the children and families whose ongoing participation made this study possible, the Quebec Institute of Statistics for data collection, and the staff of the Research Unit on Children's Psychosocial Maladjustment for their help in data access, analyses, and management.

Correspondence concerning this article should be addressed to Sylvana M Côté, University of Montreal, GRIP, 3050, Edouard–Montpetit, Montreal, QC H3T 1J7, Canada. Electronic mail may be sent to sylvana.cote1@umontreal.ca.

© 2015 The Authors
Child Development © 2015 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2015/8605-0012
DOI: 10.1111/cdev.12399

Child Care and Social Behavior: Shyness, Social Withdrawal, and Prosociality

Infants and toddlers' behavioral change during the first 6 months in child-care services presumably reflects both the initial consequences of parental unavailability and the increasing role played by peers. Infants and toddlers initially exhibit negative affect, immobility, solitary passive play with objects, and the search for adult comfort. These behaviors decrease subsequently whereas positive interactions with peers increase as children adapt to their environment (Fein, Gariboldi, & Boni, 1993). These observations suggest that children receiving nonparental child-care services during the preschool years may exhibit lower levels of shyness as compared to children who remain in parental care. Indeed, shyness is characterized by wariness and reticence in social situations (e.g., meeting unfamiliar children; Heiser, Turner, & Beidel, 2003). Shyness is associated with social withdrawal (Rubin et al., 2009), which has been conceptualized as an umbrella term describing various forms of solitary behaviors (Rubin et al., 2009). However, only two studies are consistent with the notion that preschool children who received child-care services are less shy and socially withdrawn than those who have been cared for at home by their parents (Guedeney, Foucault, Bougen, Larroque, & Mentré, 2008; Herba et al., 2013). To our knowledge, no study has yet investigated the associations between child-care services, shyness, and social withdrawal at school age. Similarly, very few studies are available regarding the effect of child care on prosociality. Although the increase in positive interactions with peers in the first months of child care (Fein et al., 1993) might suggest an association between child care and prosociality, previous studies did not detect any association (Liang, Pickles, Wood, & Simonoff, 2012; Vitaro, Desmarais-Gervais, Tremblay, & Gagnon, 1992).

Child Care and Social Behavior: Aggression and Opposition

Aggression emerges as early as the 1st year of life, is normative in early childhood, and is an important element of children's social repertoire (Nærde, Ogden, Janson, & Zachrisson, 2014; Tremblay, 2010). For example, aggression is central in the establishment of dominance relationships, which are an important feature of social groups in preschoolers (Hay & Ross, 1982; Strayer & Strayer, 1976). Throughout toddlerhood, aggression also retains a clear instrumental value in social interactions: Children can use aggression to coerce peers to achieve their aim—for instance, obtaining a coveted object (Boivin, Vitaro, et al., 2005; Hay, 2005; Hay & Ross, 1982; Patterson, Reid, & Eddy, 2002; Pingault & Blatier, 2013). As a social behavior, aggression is responsive to social clues including positive reinforcement provided by peers or adults throughout early childhood (Goldstein et al., 2001; Pingault & Blatier, 2013). Social environments like group-based child-care settings seem particularly propitious for the early expression of aggression and opposition. Aggressive and oppositional behaviors are regularly displayed in day care and children are often exposed to relevant social clues (Boivin, Vitaro, et al., 2005; Goldstein et al., 2001). Studies have indeed shown that the use of child-care services is associated with high levels of aggressive and oppositional behaviors (Averdijk et al., 2011; Belsky, 2006; NICHD Early Child Care Research Network, 2004). However, follow-up studies into middle and late childhood are crucial to fully appreciate the long-term reach of child-care experience as some effects fade away or are even reversed with time (Côté et al., 2010).

Social Group Adaptation and Fading Child-Care Effects

All children enter group-based settings at some point in their life whether it is during preschool or in elementary school years. Children with no child-care experience during the preschool years may undergo adaptation processes to the group setting when they enter kindergarten that are similar to those experienced by children who received child-care services during the preschool years. As adaptation unfolds during elementary school years, any initial difference associated with having used child care (or not) may progressively fade away. Long-term follow-ups are necessary to detect these putative fading effects after kindergarten. However, with the exception of the NICHD Study of Early Child Care and Youth Development (NICHD SECCYD), very few studies examining the role of child-care services have followed children into elementary school (Averdijk et al., 2011). The successive follow-ups of NICHD SECCYD have yielded mixed results: Using child-care services was associated with aggressive behaviors in kindergarten, but not in third grade (NICHD Early Child Care Research Network, 2004). In the follow-up between ages 4 and 12 years (Belsky et al., 2007), the initial statistically significant contribution of time in child-care services to externalizing behaviors had dissipated by 12 years, except for children in center-based services. To explain this fading difference in externalizing behaviors such as aggression between children with and without child-care experience, Belsky (2009) proposed a "diffusion" hypothesis: Child-care services lead to higher levels of aggression in children who receive these services; when all children enter kindergarten, higher aggression then diffuses to children who did not receive child-care services, resulting in converging levels of aggression (see comment by Linting & vanIJzendoorn, 2009). Alternatively, we propose that the converging levels of aggression may be explained by a social group adaptation process. Children who start to experience living in a social group—either earlier during the preschool years because they receive child-care services or later in kindergarten—undergo a transition that impacts their social behavior. Based on previous studies, we propose that children who received child-care services during the preschool years, in particular, in group settings, should initially display lower levels of shyness and social withdrawal during kindergarten, as well as higher levels of opposition and aggression than those with no such experience. Then, these initial differences in kindergarten should dissipate when children with no previous child-care experience undergo their own social group adaptation, during elementary school.

The main purpose of this study was to examine the predictive associations between child care during the preschool years and five social behaviors—shyness, social withdrawal, prosociality, opposition, and aggression—from kindergarten to the end of elementary school (Grade 6). We compared behavioral changes during the elementary school years in children with and without preschool child care experience to document the potential fading effects of child-care services. The pace at which putative initial differences in social behaviors may fade away was investigated. These differences may fade away either linearly (i.e., regular decrease) or in a nonlinear fashion (e.g., quick decrease in the first few years of elementary school). We examined the associations between the development of social behaviors and four features of preschool child-care services: (a) receiving any preschool services versus remaining in parental care, (b) the intensity of child-care services (number of hours per week), (c) the type of services (e.g., center based vs. family based), and (d) the age at which the child first received child-care services (Averdijk et al., 2011; Belsky, 2006; Jacob, 2009). Finally, we controlled for a host of confounders including children's personal characteristics, family socioeconomic characteristics, family structure, family functioning, and parenting.

The current study is correlational so no causal inference can be drawn, despite the care taken in controlling for potential confounds. The term *effect* only refers to the statistical effects of child-care predictor variables on the study outcomes and not to causal effects.

Method

Context and Participants

In the Canadian province of Québec where the study was conducted, more than 80% of children receive child-care services before they start full-time kindergarten at 5 years (Côté et al., 2007). The vast majority of children receive services in center-based child care or family-based settings, with a minority receiving individual care by a family member (e.g., grandmother) or a nanny (see below for figures in the present study). The likelihood of receiving center-based child-care services is higher for older children (Japel, Tremblay, & Côté, 2005). This is mainly due to the increasing availability of center-based spaces as the teacher-to-child ratio changes with age. In center-based care, children are of similar ages and the groups have the following teacher-to-child ratios: 1:5 between 3 months and 1½ years; 1:8 between 1½ and 4 years. In family-based settings, children vary in age, usually between 1 and 4 years, and the ratio is 1:6. In most family-based settings, two child-care workers are present (the owner of the house and an assistant) and the groups vary between 6 and 12 children. In 1997, the provincial government gradually implemented a network of publicly funded child-care services. Services were available at a low cost ($5 per day initially, now $7 per day) in family-based or center-based settings. The roll of the program was gradual, starting with 4-year-old children in 1997–1998, 3-year-old in 1999–2000, 2-year-old in 2001–2002, and 1-year-old children in 2003–2004. The families in the present study (i.e., the Québec Longitudinal Study of Child Development [QLSCD]) benefited from the low-cost spaces between 2½ and 5 years. Despite an increase in the total number of places in the Quebec child-care preschool services, the number of places in the public network was always lower than the demand, and the social selection of families into the low-cost spaces was similar to the selection operating before the public network. That is, families with more educated parents and with more financial resources succeeded in obtaining a place in the highly coveted and low-cost "early childhood centers" (Geoffroy et al., 2012; Japel et al., 2005).

QLSCD participants were drawn from the Quebec Birth Registry using stratification procedures. The resulting sample (N = 2,940) was representative of singleton live births registered in the Québec live births registry in 1997–1998. This representative sample was reduced due to nonresponse, inability to contact, or not meeting study criteria. The resulting 2,120 families constituted the initial QLSCD sample that was then surveyed annually or biennially. Details regarding sampling procedures and criteria can be found elsewhere (Jetté & Des Groseilliers, 2000). The QLSCD protocol was approved by the ethics committees of Santé Québec (Montréal, Canada), the Quebec Institute of Statistics (Montréal, Canada), and the St Justine Hospital Research Center (Montreal). Written informed consent was obtained from the primary caregiver (mostly mothers) at each data collection. Data regarding child care were collected at ages 5 months, and 1½, 2½, 3½, 4, and 5 years. Teachers rated children's social behavior five times starting from kindergarten (ages 6, 7, 8 10, 12 years). Analyses were conducted on 1,518–1,544 participants who had relevant child care and behavioral data (depending on the behavior). Table 1 shows that when compared to nonparticipants (N = 576), the 1,544 participants in the study (72.8% of the initial QLSCD sample) were less likely to be male and more likely to come from intact families with a higher socioeconomic status (SES). Their mother was also less likely to be an immigrant and more likely to be working at the time of the first assessment. The largest effect size was for SES (Hedges $g = -0.26$). A summary of sample size for each analysis is presented in the Analyses section.

Table 1
Characteristics of Participants Versus Nonparticipants

	Participants (N = 1,544) % or M (SD)	Non-participants (N = 576) % or M (SD)	Effect sizes[a]	p values
Male gender	48.6	57.3	0.078	.000
Socioeconomic status	0.06 (0.98)	−0.19 (1.04)	−0.256	.000
Work status (mother working)	19.3	12.7	0.079	.000
Immigrant status (mother immigrant)	7.6	19.4	0.170	.000
Family structure (nonintact)	17.9	22.6	0.054	.015

Note Among the nonparticipants, a subset of 220 was first excluded because they did not have relevant child-care data (see Method). These participants had the following characteristics: male gender (51.8%), socioeconomic status (−0.58 [1.06]), work status (4.1%), immigrant status (23.6%), and family structure (28.2%).
[a]Hedges g (an equivalent of Cohen d for unequal sample sizes) was used for socioeconomic status. For other variables a phicoefficient is presented.

Measures

Child Care

Every year during the preschool years, the person most knowledgeable about the child (PMK; the mother in 98% of cases) answered questions regarding whether his or her child received child-care services regularly or not as well as the type of child-care services.

The Use of Nonparental Child-Care Services

The PMK answered the following question regarding the use of child-care services: Do you currently use child care such as day care, babysitting or care by a relative or other caregiver while you (and your spouse/partner) are at work or studying? The specific formulation of the questions on child care in our study varied slightly from year to year but were consistently aiming to assess the use of a regular child-care arrangement and to distinguish it from occasional babysitting. The interviewers were trained to help parents answer about the most regular and consistent child-care arrangement. In our data set, approximately 25% of families using regular nonparental child care at each time point had a mother who was not working. This clearly indicates that child care was not only used by families with both working parents.

The variable "use of child care" distinguished between children who never received child-care services on a regular basis during the preschool years (N = 103) from those who received services at least once at any of the six measurement times (coded 1, N = 1,797). Eighty-two children were excluded because they received services only on an occasional basis (e.g., occasional babysitter). Other children were excluded because they were never rated as using child-care services regularly but classifying them as never using child-care services may have been misleading because they had at least one missing data point (N = 138). Characteristics of these children excluded based on irrelevant child-care data (total N = 220) are presented in Table 1 and are partly consistent with the fact that families who were excluded based on child-care data before the child's fifth birthday were more at risk compared to those who were excluded based on behavioral data in elementary school.

Intensity of the Nonparental Child-Care Services (Hours in Child Care)

We estimated trajectories based on the mean number of hours in child-care services per week between the ages of 5 months and 5 years. To estimate the trajectories, a nonparametric procedure, K means for longitudinal data, was used—see Genolini and Falissard (2011) for details on this procedure and Herba et al. (2013) for a recent publication using this approach. Trajectories of hours in child care are presented in Figure 1. Apart from the 103 children who never received child-care services, the three trajectories of hours in child care were as follows. First, 729 children followed a high trajectory of hours in child care, with a rapid increase from 5 months to 1½ followed by a stable trajectory between 30 and 40 hr a week from 1½ to 5 years. Second, 574 children were on a rising trajectory with regularly increasing hours in child care, almost joining the high trajectory at 4 and 5 years. Finally, a low trajectory of hours in child care grouped 494 children, with mean hours per week in child care usually below 10; however, the trajectory rose slightly at age 5 years to reach 14 weekly hours (about half the time in care compared to the rising and the high trajectories at 5 years).

Age at Entry in Child-Care Services

This variable was based on child care use at each assessment point. For instance, a child who received child-care services at the 17-month assessment but not at the 5-month assessment was assigned to the 17-month age at entry category. Note that this child may have first received child-care services at any time point after the 5-month assessment and before the 17-month assessment. Four categories were created: exclusive parental care during the preschool years (N = 103), using child-care services at the 5-month assessment (age range at assessment: 4–8 months, N = 290), first-time use at the 17-month assessment (age range at assessment = 16–19 months, N = 969), and first use at any of the following assessment times (age range at assessments = 28–68 months, N = 538).

Type of Child Care

At each assessment time, the PMK was asked about the main child-care arrangement, that is, the one used for the most hours. Most children used more than one type of main child-care arrangement between 5 months and 5 years (e.g., family based at 5 months and center care at 17 months). Because we wanted to estimate the specific contribution of the type of child care, we first restricted our analyses to children who had exclusively used one type of main child care arrangement during the pre-

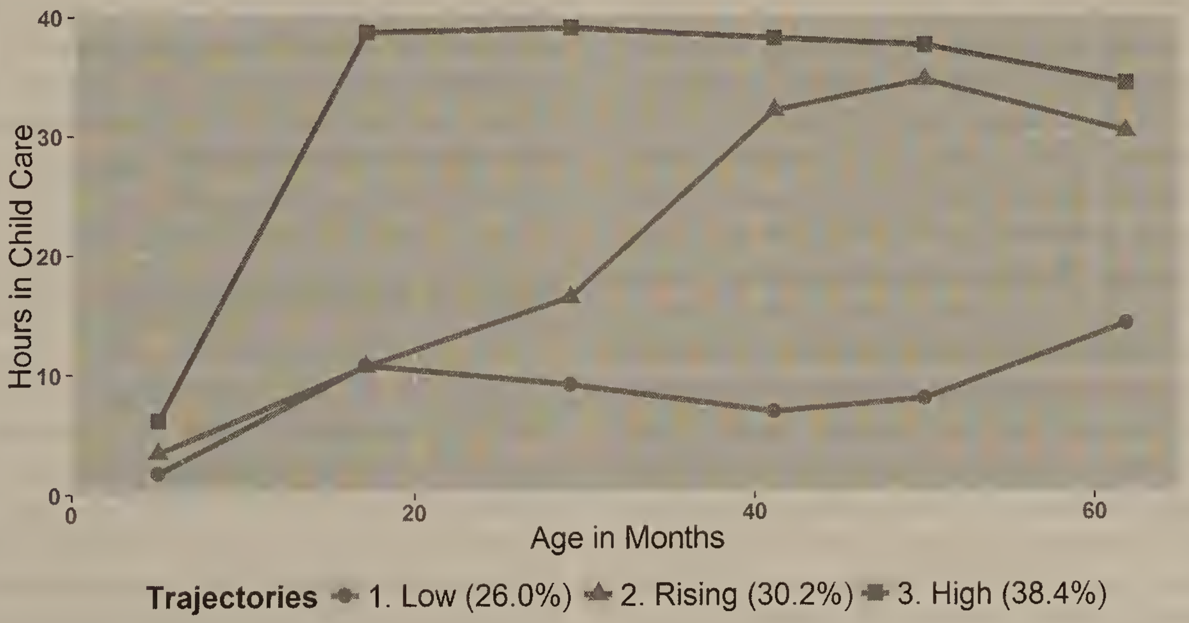


Figure 1. Trajectories of hours in child care.

school years (e.g., exclusively family based), thereby excluding those having used more than one type of child care. We allowed only one missing data point on the six time points to reduce misclassification. Children were coded as having used either center-based care (*N* = 248); family-based child care, that is, group-based child care in the home of a person unrelated to the child (*N* = 292); or individual child care (e.g., a nanny or a grandparent, *N* = 127). Dummy variables were created to compare the different types of child-care services to exclusive parental care and between themselves. Complementary analyses relaxing the conditions for missing data and for using exclusively one type of child care are briefly presented in the Complementary Analyses section in the Results.

Behavioral Outcomes

Most of the behavioral items used in QLSCD came from the Canadian National Longitudinal Study of Children and Youth (Statistics Canada, 1995), which incorporated items from the Child Behavior Checklist (Achenbach, Edelbrock, & Howell, 1987), the Ontario Child Health Study scales (Offord, Boyle, & Racine, 1989), the Preschool Behavior Questionnaire (Tremblay, Desmarais-Gervais, Gagnon, & Charlebois, 1987), and Asendorpf's social inhibition questionnaire (Asendorpf, 1993). Shyness was rated by teachers using 4 items: (a) was shy with children he/she did not know, (b) readily approached children he/she did not know (reversed item), (c) took a long time to warm up to children he/she did not know, and (d) excessively shy. Social withdrawal was assessed with 4 items: (a) tended to do things on his/her own, was rather solitary; (b) preferred to play alone rather than with other children; (c) sought the company of other children (reversed item); and (d) showed little interest for activities involving other children. Prosociality was assessed with 3 items: (a) tried to help someone who has been hurt, (b) comforted a child who was crying or upset, and (c) helped other children who were feeling sick. Opposition was assessed with 4 items: (a) was defiant or refused to comply with adults' requests or rules, (b) did not seem to feel guilty after misbehaving, (c) punishment did not change his/her behavior, and (d) had temper tantrums or hot temper. Aggression was assessed with 10 items including physical aggression items (e.g., got into fights), proactive aggression items (e.g., scared other children to get what he/she wanted), and reactive aggression (e.g., reacted in an aggressive manner when teased). All items were coded from 0 to 2 (*never or not true* = 0, *sometimes or somewhat true* = 1, and *often or very true* = 2). We computed the scores for each behavior—using mean imputation—when teachers had scored half or more of the items. For shyness however, more teachers did not score some items including "children he/she did not know." This is because teachers saw children mostly in a classroom setting where children are well acquainted. The item "excessively shy" had less missing data. To maximize the available *n*, we allowed up to three missing items for shyness. Cronbach's alphas at the five assessment points were between .69 and .77 for shyness, .67 and .73 for social withdrawal, .78 and .85 for prosociality, .80 and .84 for opposition, and .89 and .91 for aggression.

Potential Confounding Variables

Child and family socioeconomic characteristics, as well as family functioning and parenting were considered as potential confounding variables. All were assessed through PMK ratings when the child was 5 months except for information on the Cumulative Score for Neonatal Risk (CSNR; see below), which was obtained from hospital records.

Child Characteristics

Gender was coded 1 for boys, and 0 for girls. The CSNR reflected the overall health condition at birth of infants, as indexed by birth weight, length of gestation, intrauterine growth retardation, retardation of cranial perimeter growth, congenital abnormalities, APGAR score, and neonatal complications. The CSNR has been shown to be significantly associated with other indicators of health at birth, such as length of hospital stay (Séguin et al., 2001). *Difficult infant temperament* was assessed using seven items from the difficult temperament subscale of the Infant Characteristics Questionnaire (Bates, Freeland, & Lounsbury, 1979).

Family Socioeconomic Characteristics, Structure and Context

The SES of the family was derived from five variables: mother's years of schooling, spouse's years of schooling, mother's occupational status, spouse's occupational status, and household income. Occupational status was measured through the modified version of the scale (Pineo, Porter, & Mcroberts, 1977) based on Statistics Canada Standard occupational classification 1980 (Statistics

Canada, 1981). The final SES composite was standardized (Willms & Shields, 1996). Table S1 in the online Supporting Information provides additional information and descriptive statistics regarding the SES composite and the five aforementioned variables. *Work status of the mother* was coded 1 when she was working when the child was 5 months and 0 otherwise. *Immigrant status of the mother* referred to immigrant (coded 1) versus nonimmigrant (0). *Family structure* was coded 1 if the family was nonintact (i.e., reconstituted families or single-parent families) and 0 if the family was intact (i.e., the child lived with his or her two biological parents). *Presence of (at least) one sibling* (yes = 1). Two variables measured the perception of the neighborhood (Thibault, Jetté, & Desrosiers, 2001). Neighborhood safety and cohesion was measured by seven items as "safe to walk alone" and "people around here are willing to help their neighbors" (score from 0 to 4 with higher scores indicating lower safety). The *social problem* measure was a composite of six items, including drug selling or use or excessive drinking in public in the neighborhood (score from 0 to 4 with *lower* scores indicating more problems).

Family Functioning and Parenting

Family functioning was measured at 5 months using 12 items concerning communication, problem resolution, expression of affection, and control of disruptive behavior (Statistics Canada, 1995). Parenting was measured with four dimensions of the parental Cognitions and Conduct Toward the Infant Scale (Boivin et al., 2005): perceived parental self-efficacy (4 items), perceived parental impact (5 items), hostile-reactive parenting (3 items), and overprotection (4 items). Items from the Infant-Toddler Home Observation for Measurement of the Environment (Caldwell & Bradley, 1984) were used to assess emotional and verbal responsivity of the mother (10 items) and maternal involvement with the child (5 items). Both scores ranged from 0 to 10; higher scores indicate a better environment. Finally, maternal depressive symptoms were assessed using a short version of the Center for Epidemiologic Studies Depression Scale (Radloff, 1977).

Analyses

Sample Size for Analyses

Among the 1,900 children with relevant child-care data, 1,544 had at least one assessment for each of shyness, social withdrawal, and opposition between 6 and 12 years and were thus included in analyses with and without covariates. A subset of 1,543 of the 1,544 participants were included for aggression and a subset of 1,518 out of the 1,544 participants were included for prosociality. Analyses regarding type of child-care services included: 640 participants for shyness, social withdrawal, opposition, and aggression, and 633 for prosociality (see Type of Child Care in the Measures section).

Statistics

Latent growth curves were used to model the contribution of child-care services to the initial levels (at 6 years) and the change (from 6 to 12 years) of raw scores of shyness, social withdrawal, opposition and aggression and prosociality. For each behavior, we first estimated a baseline model without covariates and with child care use as the only predictor, which allowed us to model the behavioral development of users versus nonusers. In particular, we examined the possibility of nonlinear behavioral development in the two groups. Once an adequately fitting baseline model was obtained for each behavior, we estimated a model controlling for the contributions of all confounding variables to the intercept and growth parameter(s). We did not exclude control variables based on bivariate relations as it may mask multivariate effects (Linting & van IJzendoorn, 2009). The models without and with covariates were estimated for each of the four child-care variables, that is, use, intensity, age at entry and type of child care. Full information maximum likelihood (ML) was used to handle missing data. An ML estimator can be used for values of skewness under 2 and kurtosis under 7 (Curran, West, & Finch, 1996). Shyness, social withdrawal, prosociality, and opposition satisfied these conditions at all measurement points. Aggression kurtosis was also under 7 at all times, but skewness was above 2 at four time points (maximum 2.26). Therefore, an ML robust estimator, with robust standard errors and an adjusted chi-square, was used for aggression. The model's fit was assessed using the model chi-square: A nonsignificant model chi-square indicates a good fit, in particular, in larger samples. We also report three approximate fit indexes (Kline, 2011): comparative fit index (CFI) for which values close to 1 indicate a better fit, root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR); for the two latter, values close to 0 indicate better fit. No consensus exists on cutoff values for these indexes, but values close to 0.95 for CFI, 0.06 for RMSEA, and 0.08 for SRMR have been suggested to conclude that there is a relatively good fit between the model and the data (Hu & Bentler, 1999). Kline (2011) recommended to never ignore the possibility of specification error indicated by a significant chi-square even when approximate fit indices look favorable and to check that each correlation residual is under 0.10. Correlation residuals indicate the discrepancy between the model predicted correlations and the observed correlations. Effect sizes were calculated following Hancock's (2001) approach for latent variables (e.g., the unstandardized estimate of the effect of the binary predictor on the intercept is divided by the square root of the residual variance of the intercept). Analyses were conducted with R 3.0.1, using the package lavaan 0.5-13 to fit the growth curves (Rosseel, 2012), and the package Kml 2.2 to estimate trajectories of hours in child care (Genolini & Falissard, 2011).

Complementary Analyses

Sensitivity Analysis

Shyness, social withdrawal, prosociality, opposition, and aggression were not assessed at 5 months, which precludes a control for baseline levels of the target behaviors before entry into child care. This lack of control might introduce a bias due to a child effect (McCartney et al., 2010); for example, shy and withdrawn children may react to attempts to place them in child care, leading parents to keep them at home. If this was the case, any group differences in kindergarten could reflect initial differences maintained over time rather than the effect of child care. Controlling for the first behavioral assessment available (PMK ratings at the 17-month assessment) is also problematic as the majority of participating children had received child-care services by 17 months. Consequently, behavioral differences at that age could already be due to exposure to child-care services and controlling for these behaviors would unduly underestimate the contribution of child care (McCartney et al., 2010). Potential confounders should be measured before the exposure to the predictor of interest, that is, child-care services in the present case (Ho, Imai, King, & Stuart, 2007). Therefore, although we did not include the 17-month behavioral assessments in the main analyses, we reran the models with covariates measuring the contribution of child care use including the 17-month assessment as covariates (e.g., 17-month shyness in the shyness model with covariates). Shyness, opposition, prosociality, and physical aggression (used as proxy for later aggression) were rated by the PMK at age 17 months. However, social withdrawal was not available. Additional complementary analyses are reported in the Complementary Analyses section in the Results.

Propensity Score Matching

To test whether the results were sensitive to the use of regression analyses to adjust for covariates, we reanalyzed the statistically significant results using propensity score matching. In brief, the analyses involved three steps using three additional R packages. First, 50 data sets were imputed using package mice 2.22 (Van Buuren & Groothuis-Oudshoorn, 2011). Second, propensity score matching was implemented on each data set using package MatchIt 2.4-21 (Ho, Imai, King, & Stuart, 2011). Matching was conducted with all covariates as well as with the four behavioral scores assessed at age 17 months (see Sensitivity Analysis above). Third, the weights obtained from the matching procedure were used to reanalyze the latent growth models using package lavaan.survey 1.1.1 (Oberski, 2014). This package combines the weighted latent growth model estimates across the imputed data sets and implements robust standard errors to account for the use of weights. Procedures are described in detail in the online Supporting Information.

Results

Fit indices for all models are shown in Table S2. Detailed results are provided in Table 2 for each model. In Table 2, the contribution of each child-care variable (i.e., use, intensity, age at entry, type of services) has been estimated separately, controlling for all confounds. Complementary analyses including several child-care variables are reported below. Here, we provide parameter estimates, 95% confidence intervals, *p* values, and standardized effect sizes for the statistically significant findings.

Shyness

We first fitted the baseline model without covariates, with child care use as the sole predictor. A model with a linear slope fitted the data well as shown by a nonsignificant chi-square and adequate fit indices. Figure 2a shows the close correspondence between model-based predicted means and

Table 2

Contribution of Nonmaternal Child Care to Childhood Behaviors, Controlling for Covariates

		Intercept				Slope			
	Child care[a]	Estimate	95% CI lower	95% CI upper	*p* value	Estimate	95% CI lower	95% CI upper	*p* value
Shyness	Use	−0.80	−1.18	−0.41	0.000	0.07	−0.03	0.16	0.166
	Hours: low versus no	−0.56	−0.97	−0.15	0.008	0.04	−0.06	0.14	0.418
	Hours: rising versus no	−0.91	−1.32	−0.51	0.000	0.08	−0.02	0.18	0.113
	Hours: high versus no	−0.94	−1.35	−0.52	0.000	0.08	−0.02	0.18	0.116
	Hours: rising versus low	−0.35	−0.60	−0.10	0.006	0.04	−0.02	0.10	0.200
	Hours: high versus low	−0.38	−0.62	−0.13	0.003	0.04	−0.02	0.10	0.193
	Entry: 5 m versus no	−0.67	−1.18	−0.16	0.010	0.00	−0.12	0.12	0.946
	Entry: 17 m versus no	−0.87	−1.27	−0.47	0.000	0.07	−0.02	0.17	0.140
	Entry: 29+ m versus no	−0.75	−1.15	−0.34	0.000	0.07	−0.03	0.17	0.153
	Entry: 5 m versus 17 m	0.20	−0.16	0.55	0.276	−0.07	−0.15	0.01	0.100
	Entry: 29+ m versus 17	0.13	−0.11	0.36	0.290	0.00	−0.06	0.05	0.960
	Type: family based versus no	−0.72	−1.16	−0.28	0.001	0.05	−0.06	0.15	0.360
	Type: center versus no	−0.68	−1.13	−0.23	0.003	0.03	−0.08	0.14	0.610
	Type: individual versus no	−0.57	−1.08	−0.06	0.030	0.08	−0.04	0.21	0.205
	Type: family based versus center	−0.05	−0.39	0.30	0.800	0.02	−0.06	0.10	0.645
	Type: individual versus center	0.11	−0.32	0.54	0.621	0.05	−0.05	0.16	0.339
Social withdrawal	Use	−0.60	−0.95	−0.26	0.001	0.07	−0.02	0.17	0.124
	Hours: low versus no	−0.46	−0.82	−0.09	0.014	0.06	−0.04	0.16	0.263
	Hours: rising versus no	−0.69	−1.06	−0.33	0.000	0.10	0.00	0.19	0.058
	Hours: high versus no	−0.66	−1.03	−0.30	0.000	0.07	−0.03	0.17	0.197
	Hours: rising versus low	−0.24	−0.46	−0.02	0.036	0.04	−0.02	0.10	0.199
	Hours: high versus low	−0.20	−0.42	0.02	0.069	0.01	−0.05	0.07	0.764
	Entry: 5 m versus no	−0.45	−0.90	0.00	0.051	0.01	−0.11	0.13	0.910
	Entry: 17 m versus no	−0.63	−0.99	−0.28	0.000	0.05	−0.05	0.14	0.343
	Entry: 29+ m versus no	−0.60	−0.96	−0.24	0.001	0.12	0.02	0.21	0.020
	Entry: 5 m versus 17 m	0.19	−0.13	0.50	0.244	−0.04	−0.12	0.04	0.337
	Entry: 29+ m versus 17	0.04	−0.17	0.24	0.736	0.07	0.01	0.12	0.013
	Type: family based versus no	−0.57	−0.97	−0.18	0.005	0.04	−0.06	0.15	0.428
	Type: center versus no	−0.64	−1.05	−0.23	0.002	0.08	−0.03	0.19	0.157
	Type: individual versus no	−0.75	−1.22	−0.29	0.001	0.12	0.00	0.25	0.048
	Type: family based versus center	0.06	−0.25	0.38	0.692	−0.04	−0.12	0.05	0.385
	Type: individual versus center	−0.12	−0.51	0.27	0.561	0.06	−0.06	0.15	0.385
Prosociality	Use	0.13	−0.19	0.44	0.427	−0.02	−0.11	0.08	0.731
	Hours: low versus no	0.18	−0.16	0.51	0.303	−0.04	−0.14	0.07	0.485
	Hours: rising versus no	0.12	−0.21	0.46	0.470	0.00	−0.10	0.10	1.000
	Hours: high versus no	0.07	−0.27	0.41	0.680	−0.01	−0.12	0.09	0.805
	Hours: rising versus low	−0.05	−0.26	0.16	0.616	0.04	−0.03	0.10	0.276
	Hours: high versus low	−0.11	−0.31	0.10	0.321	0.02	−0.04	0.09	0.473
	Entry: 5 m versus no	−0.02	−0.44	0.40	0.909	0.04	−0.09	0.16	0.571
	Entry: 17 m versus no	0.13	−0.20	0.45	0.440	−0.02	−0.12	0.08	0.682
	Entry: 29+ m versus no	0.15	−0.18	0.48	0.371	−0.02	−0.12	0.08	0.731
	Entry: 5 m versus 17 m	−0.15	−0.45	0.15	0.316	0.06	−0.03	0.15	0.210
	Entry: 29+ m versus 17	0.02	−0.17	0.22	0.820	0.00	−0.06	0.06	0.920
	Type: family based versus no	0.07	−0.30	0.44	0.713	0.00	−0.12	0.11	0.983
	Type: center versus no	0.21	−0.17	0.59	0.285	0.01	−0.13	0.11	0.870
	Type: individual versus no	0.17	−0.27	0.60	0.446	−0.02	−0.16	0.12	0.795
	Type: family based versus center	−0.14	−0.43	0.16	0.373	0.00	−0.10	0.09	0.979
	Type: individual versus center	−0.04	−0.41	0.33	0.835	−0.02	−0.14	0.11	0.790

Table 2
Continued

		Intercept				Slope			
	Child care[a]	Estimate	95% CI lower	95% CI upper	p value	Estimate	95% CI lower	95% CI upper	p value
Opposition	Use	0.70	0.30	1.11	0.001	−0.30	−0.51	−0.08	0.007
	Hours: low versus no	0.50	0.07	0.94	0.022	−0.26	−0.49	−0.04	0.023
	Hours: rising versus no	0.76	0.33	1.18	0.001	−0.30	−0.53	−0.08	0.009
	Hours: high versus no	0.89	0.46	1.32	0.000	−0.33	−0.56	−0.11	0.004
	Hours: rising versus low	0.25	−0.01	0.52	0.061	−0.04	−0.18	0.10	0.597
	Hours: high versus low	0.39	0.12	0.65	0.004	−0.07	−0.21	0.07	0.329
	Entry: 5 m versus no	0.70	0.17	1.24	0.010	−0.27	−0.55	0.01	0.060
	Entry: 17 m versus no	0.79	0.36	1.21	0.000	−0.33	−0.56	−0.11	0.003
	Entry: 29+ m versus no	0.61	0.19	1.04	0.005	−0.26	−0.48	−0.03	0.024
	Entry: 5 m versus 17 m	−0.08	−0.46	0.30	0.670	0.07	−0.13	0.26	0.510
	Entry: 29+ m versus 17	−0.17	−0.42	0.07	0.165	0.08	−0.05	0.20	0.244
	Type: family based versus no	0.55	0.08	1.02	0.023	−0.28	−0.52	−0.03	0.027
	Type: center versus no	0.84	0.34	1.32	0.001	−0.35	−0.61	−0.09	0.008
	Type: individual versus no	0.30	−0.24	0.85	0.277	−0.24	−0.53	0.05	0.099
	Type: family based versus center	−0.29	−0.66	0.09	0.129	0.07	−0.12	0.27	0.452
	Type: individual versus center	−0.53	−1.00	−0.07	0.024	0.11	−0.13	0.35	0.367
Aggression	Use	1.65	0.96	2.35	0.000	−0.61	−1.09	−0.13	0.012
	Hours: low versus no	1.25	0.49	2.01	0.001	−0.56	−0.98	−0.14	0.008
	Hours: rising versus no	1.87	1.09	2.65	0.000	−0.67	−1.10	−0.25	0.002
	Hours: high versus no	1.87	1.10	2.64	0.000	−0.65	−1.06	−0.23	0.002
	Hours: rising versus low	0.62	0.05	1.18	0.034	−0.11	−0.39	0.17	0.428
	Hours: high versus low	0.62	0.08	1.16	0.025	−0.09	−0.36	0.18	0.523
	Entry: 5 m versus no	1.99	0.98	3.00	0.000	−0.74	−1.35	−0.13	0.018
	Entry: 17 m versus no	1.84	1.10	2.57	0.000	−0.63	−1.13	−0.13	0.014
	Entry: 29+ m versus no	1.41	0.65	2.16	0.000	−0.58	−1.09	−0.07	0.025
	Entry: 5 m versus 17 m	0.15	−0.65	0.96	0.708	−0.11	−0.55	0.32	0.605
	Entry: 29+ m versus 17	−0.43	−0.96	0.10	0.111	0.05	−0.25	0.35	0.755
	Type: family based versus no	1.25	0.42	2.07	0.003	−0.61	−1.12	−0.09	0.021
	Type: center versus no	2.02	1.10	2.94	0.000	−0.74	−1.27	−0.21	0.006
	Type: individual versus no	0.71	−0.22	1.64	0.139	−0.36	−0.98	0.25	0.246
	Type: family based versus center	−0.78	−1.55	0.00	0.050	0.13	−0.28	0.55	0.530
	Type: individual versus center	−1.31	−2.18	−0.45	0.008	0.38	−0.13	0.89	0.144

Note. The table presents the contribution of each child-care variable to the intercept and the slope of shyness, social withdrawal, prosociality, opposition, and aggression. The contribution of each child-care variable was estimated in separate models (e.g., the age at entry model did not include the type variable). All models controlled for the contributions of all confounding variables to the intercept and the slope. m = month.

[a] In this column, "Use" refers to the contribution of having received nonparental child-care services at least once between 5 and 60 months, contrasted with having remained in exclusive parental care. Contrasts are mentioned in each row for "Hours," "Entry," and "Type." For instance, "Entry: 29+ m versus 17 m" means entering in child care at the 29 months assessment versus entering at the 17 months assessment. In addition to the contribution of each child-care category the table shows additional pairwise contrasts. For instance, regarding type of child care, three main estimates are presented for: (a) family-based versus no child care, (b) center versus no child care, and (c) individual versus no child-care services (i.e., exclusive parental care). Two additional contrasts are also presented: family based versus center and individual versus center. Note that the latter two estimates do not correspond to additional predictors but simply to additional contrasts between existing categories.

the observed means from 6 to 12 years. Children having used child care at least once between 5 months and 5 years were less likely to be seen as shy in kindergarten. Table 2 presents the results (controlling for the contributions of all covariates on the intercept and on the slope) and shows that the contribution of child care use to the intercept was statistically significant (−0.80, 95% CI [−1.18, −0.41], $p < 0.001$, $d = .62$), whereas the positive contribution to the slope was not. Table S3 shows the contribution of all covariates to childhood levels of shyness.

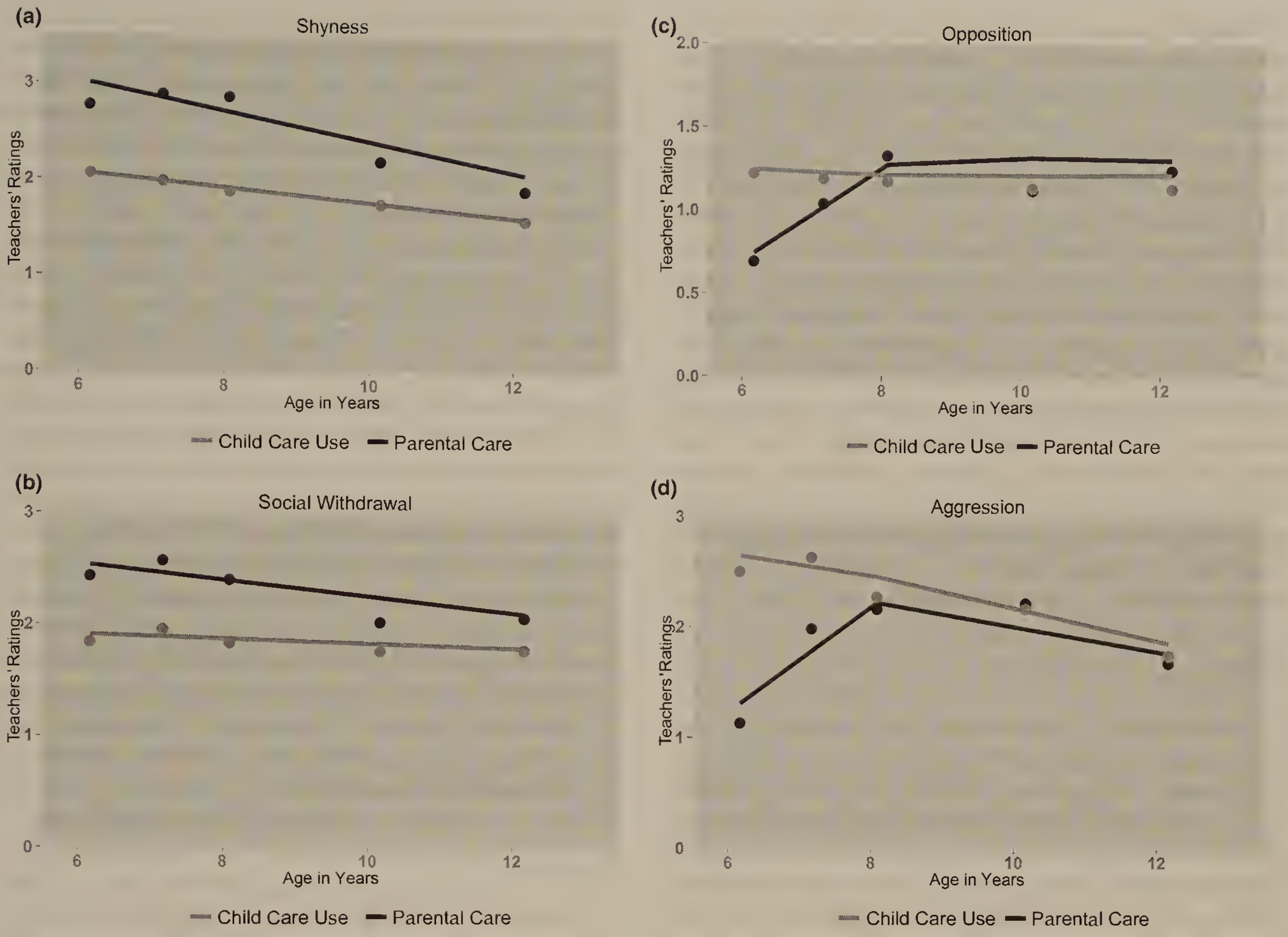


Figure 2. Social behaviors from 6 to 12 years according to early child-care use. (a) Shyness. (b) Social withdrawal. (c) Opposition. (d) Aggression. Fitted mean values (lines) based on models without covariates and observed mean values (points) were plotted for two groups: children receiving child-care services and children in exclusive parental care. Contributions of the use of child-care services to the intercepts were statistically significant for all four behaviors. Statistically significant contributions to the slopes were detected for opposition and aggression but not for shyness and social withdrawal (although they reached significance in propensity score matching analyses; see Table S10 in the online Supporting Information). Note that the four graphs represent the baseline models for the variable use of child-care services. Refer to the text for a finer understanding of the results for other variables (e.g., the trajectory of aggression for children who received child-care services depend on the type of child care).

With respect to intensity, children in all trajectories had statistically significant lower initial shyness levels when compared to children in exclusive parental care: low trajectory (−0.56, 95% CI [−0.97, −0.15], $p = 0.008$, $d = .44$), rising (−0.91, 95% CI [−1.32, −0.51], $p < 0.001$, $d = .72$), and high (−0.94, 95% CI [−1.35, −0.52], $p < 0.001$, $d = .74$). In addition, children in the rising and high trajectories had similar levels of shyness, both being statistically significantly lower than levels for children in the low trajectory. Compared to children who never used preschool child-care services, those who did were initially less shy irrespective of their age at entry. All children also had lower initial levels of shyness irrespective of the type of child-care services they received (see Table 2).

To evaluate if the differences in shyness would be maintained at age 12 years, the same models were tested again with the intercept at 12 years. Children in the high trajectory of hours in child-care (−0.46, 95% CI [−0.92, −0.01], $p = 0.047$, $d = .49$) as well as children with age at entry at 5 months (−0.65, 95% CI [−1.20, −0.01], $p = 0.02$, $d = .69$) were still statistically significantly less shy at age 12 years.

Social Withdrawal

As for shyness, a linear model fitted the data adequately (see Figure 2b). In analyses with covariates (see Table 2), child-care use predicted lower initial social withdrawal (−0.60, 95% CI [−0.95, −0.26], $p = 0.001$, $d = .60$), but the contribution to the slope was not statistically significant. Table S4 shows the contribution of all covariates to childhood levels of social withdrawal. Children in the low (−0.46, 95% CI [−0.82, −0.09], $p = 0.014$, $d = .46$), rising (−0.69, 95% CI [−1.06, −0.33], $p < 0.001$, $d = .70$), and high trajectories (−0.66, 95% CI [−1.03, −0.30], $p < 0.001$, $d = .66$) of hours in child-care services had lower initial levels of social withdrawal. In addition, children in the rising child-care hours trajectory had statistically significantly lower levels of social withdrawal than children in the low trajectory. Note that a similar pattern was observed for the high trajectory but did not reach statistical significance at conventional values. All children having used child-care services during the preschool years had lower initial levels of social withdrawal irrespective of age at entry or type of services received (see Table 2). None of these associations were maintained at 12 years.

Prosociality

We retained a linear model for prosociality (additional comments on model fitting are presented in the online Supporting Information). In analyses with covariates, no contribution was detected whether for use, intensity, age at entry, or type of child care (see Table 2). Table S5 shows the contribution of all covariates to childhood prosociality.

Opposition

As illustrated in Figure 2c, children who never used child-care services before age 5 years started with lower levels of opposition in kindergarten but rapidly caught up with children who had used child-care services. The two groups had then virtually identical flat levels of opposition from age 8 years onward. We freed the last two loadings to capture this nonlinear change. Figure 2c shows the correspondence between fitted and observed means, indicating that this model effectively captured the nonlinear change in the nonusers (additional comments are provided online).

In analyses with covariates (see Table 2), children having used child-care services had higher levels of opposition at 6 years (0.70, 95% CI [0.30, 1.11], $p = 0.001$, $d = .44$) and their slope from 6 to 8 years was statistically significantly less sharp compared to children who never received nonparental child-care services. Table S6 displays the contribution of all covariates.

The pattern of results displayed in Figure 2c was observed irrespective of age at entry. Furthermore, this pattern of results was observed only for family-based and center-based services and not for individual child-care services. Specifically, children in family-based and center-based services had statistically significantly higher levels of opposition than those in individual care but did not differ between themselves (see Table 2). Regarding hours in child care, children in the low (0.50, 95% CI [0.07, 0.94], $p = 0.022$, $d = .32$), rising (0.76, 95% CI [0.33, 1.18], $p = 0.001$, $d = .48$), and high (0.89, 95% CI [0.46, 1.32], $p < 0.001$, $d = .57$) trajectories statistically significantly differed from the children in exclusive parental care for both the intercept and the slope. In addition, consistent with a dose effect, children in the high trajectory had statistically significantly higher initial levels of opposition, when compared to children in the low trajectory.

As seen in Figure 2c, children who used child-care services during the preschool years had similar levels of opposition to those in exclusive parental care at 8 years. We reassessed the aforementioned models by setting the intercept at age 8 years to see if differences observed at age 6 years remained statistically significant. There was no more statistically significant difference at 8.

Aggression

Figure 2d shows a change in aggression similar to that found for opposition for children who did not receive preschool child-care services. This change could not be summarized with a linear slope. We used a piecewise growth model, which included two linear slopes: The first modeled the change from 6 to 8 years (three time points) and the second captured the change from 8 to 12 years. Figure 2d shows that predicted means were close to observed means in this model, which was not the case in the previous models.

Compared to children who did not, children who used child-care services had statistically significantly higher initial levels of aggression (1.65, 95% CI [0.96, 2.45], $p < 0.001$, $d = .52$). Their slope from ages 6 to 8 years also differed. As illustrated in Figure 2d, from ages 8 to 12 years, the change is very similar in the two groups and no contribution to the second slope was detected. To summarize, children who did not receive child-care services during preschool started kindergarten with lower levels of aggression but then caught up rapidly until age 8 years when they followed the same course as children who received child-care services during the preschool years. Table S7 shows the contributions of all covariates to childhood levels of aggression.

Although children who received child-care services earlier (at 5 and 17 months) had higher initial levels of aggression compared to those who received the services later (i.e., at 30 months or after), the differences were not statistically significant. Children in the low, rising, and high trajectories of hours in preschool child-care services all differed from children who did not use the services, with effects similar to those displayed in Figure 2d: low trajectory (1.25, 95% CI [0.49, 2.01], $p = 0.001$, $d = .40$), rising trajectory (1.87, 95% CI [1.09, 2.65], $p < 0.001$, $d = .58$), and high trajectory (1.87, 95% CI [1.10, 2.64], $p < 0.001$, $d = .58$). In addition, children in both the rising and high trajectories had higher initial levels of aggression compared to children in the low trajectory. Regarding the type of child care, patterns similar to those displayed in Figure 2d were found for those receiving center-based services and for those receiving family-based services but not for those receiving individual services. In particular, initial levels of aggression were higher in children in center-based services (2.02, 95% CI [1.10, 2.94], $p < 0.001$, $d = .62$) and family-based services (1.25, 95% CI [0.42, 2.07], $p = 0.003$, $d = .38$), but not for those using individual services. In addition, the difference between children in center-based care and individual care was statistically significant. No difference in aggression levels remained statistically significant at 8 years.

Complementary Analyses

Sensitivity Analysis

We controlled for the 17-month behavioral measures in models with covariates assessing the contribution of child-care use (e.g., controlling for shyness at 17 months in the shyness model with covariates). PMK ratings of shyness, opposition, and physical aggression (but not prosociality) at 17 months were statistically significantly associated with their corresponding behavior at age 6 years (i.e., with the intercept of the model with covariates). All statistically significant contributions of child-care use to the initial level or to the slope in the models with covariates (see Table 2) remained statistically significant after controlling for behavior at 17 months and none of the estimates decreased by more than 6.3%.

Type of Child-Care Services

To be more inclusive in the analyses regarding the type of child-care services, we considered combinations of main child-care arrangements: (a) group care only, with children exclusively ingroup settings, that is, during the preschool years they used both center and family-based services; (b) other combinations, with children who received a mix of individual services and services in a group setting (i.e., attended individual care and center and/or family-based services). Results were: (a) very similar to the results presented above for children using exclusively one type of main child-care arrangement across the preschool years, despite relaxing the missing data condition and (b) estimates for combination of types were consistent with estimates for the exclusive types (e.g., for the family-based/center combination the estimate was between the estimates for the pure types family-based and center-based care). The only exception was that children in center-based services had statistically significantly higher initial levels of opposition and aggression not only when compared to children who did not receive preschool child-care services and children in individual care, as in the main analyses, but also, to a lesser extent, when compared to children in family-based services. Results are further detailed in Table S8.

Multiple Child-Care Variables

As stated above, the contribution of each child-care variable was estimated separately (see Table 2). For shyness and social withdrawal, child-care type and age at entry did not contribute, leaving only the statistically significant contributions of use of child-care services and intensity. Because the intensity variable (trajectories of hours in child care) is a specification of the binary use of child-care services variable, including them in the same model is not informative. For opposition and aggression, however, the type of child care services and intensity of child care made a statistically significant contribution. Therefore, the two variables were included together to test whether children in center-based services still had elevated levels of aggression when the intensity of care was controlled for. The results were essentially unchanged for type of care. In addition, children in the high trajectory of hours in child-care services still had statistically significantly higher initial levels of opposition compared to children in the low trajectory. The same pattern was observed for aggression but it did not reach statistical significance at conventional values. Detailed results are presented in Table S9.

Propensity Score Matching

Results are briefly described below and detailed in the online Supporting Information.. Three sets of analyses were conducted. First, results regarding the contribution of child-care use were reanalyzed by matching children who did not use child-care services with those who did. To assess matching efficiency, the mean differences between the two groups for all covariates before and after matching are presented in Table S10. Differences between the two groups were drastically reduced after matching. Final analyses are presented in Table S11 and results were very close to those obtained with covariate-adjusted analyses. Next, the contribution of trajectories of hours in child-care services was reanalyzed. Because the pattern of results was very similar for the rising and the high hour trajectories (see Table 2), we compared children in the low trajectory with those in the rising/high trajectory (see Table S12). Estimates were similar to those resulting from regression analyses (Table S13) and were statistically significant for shyness and social withdrawal, and marginally significant for opposition and aggression. Finally, we tested whether the contribution of the type of child-care services to opposition and aggression was maintained when using propensity score matching analysis. Specifically, we tested whether children in center-based services still differed from children in individual services. Results are presented in Tables S14 and S15; estimates were very similar to those found in regression analyses and statistically significant for both aggression and opposition. Overall, results were remarkably similar when using either: (a) full information ML to deal with missing data and regression analyses to control for covariates in the main analyses or (2) multivariate imputation by chained equation and propensity score matching in the complementary analyses, which we briefly discuss in the online Supporting Information.

Discussion

The purpose of this study was to examine the predictive association between the use of regular nonparental child-care services and social behaviors in kindergarten and elementary school. Compared to those who remained in parental care throughout the preschool years, children who received child-care services at least once between 5 months and 5 years were rated by their kindergarten teacher as less shy, less socially withdrawn, more oppositional, and more aggressive. However, these initial differences dissipated with age. We first discuss the results for each behavior before considering alternative developmental hypotheses to explain the dissipating association over time.

Shyness and Social Withdrawal

Compared to children who remained in exclusive parental care, children who received child-care services during the preschool years were seen by kindergarten teachers as less shy and less socially withdrawn. Consistent with a dose effect, children spending more hours in child-care services were also less shy and less socially withdrawn than children spending fewer hours in child care. However, most of these differences had disappeared by the end of elementary school (i.e., Grade 6, age 12 years). Shy children tend to avoid activities and social events (Kingery, Erdley, Marshall, Whitaker, & Reuter, 2010). Child-care services can create a context where shy children have to deal with their social wariness early in life. Therefore, child-care services may work in a similar fashion to intervention programs that provide support for shy children in progressively overcoming their social wariness rather than avoiding social situations (Rapee, Schniering & Hudson, 2009). There is also evidence that children's behaviors change substantially in the first 6 months of child care, from frequent displays of negative affect and comfort contact with adults toward more positive interactions with peers (Fein et al., 1993). This suggests that children, reluctant at first, get progressively acquainted to a new type of interaction in a nonfamilial environment. As a result, children having experienced child care are likely better prepared for social interactions and may be seen by their kindergarten teachers as less shy and socially withdrawn than those having never experienced child-care services.

Opposition and Aggression

The patterns found for opposition and aggression were quite different from those found for shyness and social withdrawal. Children who received child-care services, compared to those who remained in parental care, displayed higher levels of opposition and aggression, especially those who were spending more hours in child-care services. However, children not exposed to child-care services steadily caught up with those who were, reaching the levels of aggression and opposition of children with child-care experience around age 8 years. For both opposition and aggression, the contributions were statistically significant for children in center-based care and those using family-based services; however, this was not the case for children in individual child care. Overall, these results are consistent with studies that have reported a link between child-care services and aggression and defiant behaviors in kindergarten (e.g., Averdijk et al., 2011; NICHD Early Child Care Research Network, 2004). Early childhood is an important period for the expression and development of aggression (Tremblay, 2010). Children receiving child-care services have more opportunities to experience conflict with peers and adults and express aggressive and oppositional behaviors. This is consistent with the fact that we found higher levels of aggression and opposition for children in both types of group-based child-care services (family based and center based) but not for children in individual services. Averdijk et al. (2011) reported similar results and McCartney et al. (2010) suggested that exposition to larger groups of peers was a potential process explaining the association between child care and higher externalizing behaviors.

Selection, Dose Effect, and Causality

Given the correlational nature of the present study, no definite claim about causality can be made. However, in nonexperimental studies, elements consistent or inconsistent with causality can be pointed out (Imai, King, & Stuart, 2008), in particular, threats to the plausibility of causality. One such potential threat is selection effects. Different methods have been used to account for selection effects in the child-care literature (e.g., regression with covariates or propensity score techniques) and variations in methods have led to more or less conservative estimates of the contribution of child-care services. However, in the present study, regression and propensity score matching techniques to adjust for potential confounds yielded very similar results. Furthermore, it would be expected that any social selection effect remaining after statistical adjustments would play in favor of children from high-SES families showing less behavior problems in kindergarten. Instead, our results showed more initial problems with aggression and opposition in children who received child-care services (more likely of higher SES families) than in children who did not receive child-care services (more likely of lower SES families). One could argue that a “child effect,” for example, preexisting temperamental or behavioral differences influencing whether a child receives services or not, could also play out in selection. However, consistent with a previous study that found no such effect on externalizing behaviors (McCartney et al., 2010), the sensitivity analysis suggested that this type of effect was not likely in the present study. Finally, our main finding is that the positive (for shyness and social withdrawal) or negative effects (for opposition and aggression) of child care fade out during the elementary school years, because children who did not receive child-care services catch up with those who did. Such effects are not compatible with social selection or child effects, as we would expect such selection processes to lead to enduring differences between these two groups of children. Misidentification is another potential threat to the inference of causality. For example, in the present study, we cannot rule out the possibility that the results might be a function of having working parents as it is of receiving child-care services. Finally, McCartney et al. (2010) have argued that, to be consistent with causality, a dose–response relation between time spent in child care and outcomes should be observed (the alternative being that any exposure to child care at any developmental stage produces similar outcomes). In the present study, hours spent in child care showed signs of a dose–response effect.

A Social Group Adaptation Hypothesis

We first propose to interpret the apparent dissipation effect from the perspective of a social group adaptation hypothesis and then contrast this hypothesis with the diffusion hypothesis proposed by Belsky (2009) for externalizing behavior problems.

Whether in early childhood or later, children integrate into social group settings (child care, kindergarten, elementary school) and have to adapt their social behavior accordingly. In the case of early entry into a group-based child-care setting, living in a social group is generally new for all children in the group. In such a context, children initially tend to manifest shy and withdrawn behaviors. As they become familiar with the social group, they progressively become less shy and less socially withdrawn. At the same time, the social group offers opportunities to engage in aggressive and oppositional behaviors due to the increased social interactions, the increased likelihood of conflict, and the rules set by teachers. When children first integrate a social group later, for instance, in kindergarten, the social setting is new for them but generally not new for most children in the group. The experience of children entering late into a group-based setting is thus different from those who enter early: The former are confronted with a preexisting context; that is, they are entering a group where most children have already adapted to living in a social group. As such, not only do these children have to adapt to a group situation per se, but they have to adapt to a group situation with experienced children. In this context the adaptation to the group may take the form of a "regression to the group mean." This was particularly striking in the case of aggression (as illustrated in Figure 2d). Specifically, children not exposed to child-care services before kindergarten (i.e., they remained in parental care) started at lower levels of aggression in kindergarten and steadily increased until they caught up, at around age 8 years, with children who had received child-care services. The growth curve then took a sharp turn and started to decrease, mimicking the change in children with previous child-care experience. From this "social group adaptation hypothesis," the presumed behavioral consequences of child-care services might be better conceived as resulting from an adaptation to social group settings rather than as specific effects of child care. Whether this adaptation occurs in early childhood (i.e., child care) or at school entry (i.e., kindergarten), may not be as relevant: Children with no preschool child-care experience seemed to have completed their own social transition by age 12 years and were largely similar to children with child-care experience.

Belsky (2009) suggested that high levels of aggression in children having received child-care services do not dissipate later on but that, instead, levels of aggression in children without child-care experience increase upon contact with children who have child-care experience. In other words, the adverse effects of child care on children's levels of aggression diffuse to children without child-care history when they enter kindergarten (Belsky, 2009; Linting & van IJzendoorn, 2009). Our results are partly consistent with this proposition in that during elementary school, children who had not used child-care services during the preschool years caught up with those who had used child-care services (see Figure 2d). However, our study shows that any such diffusion effect could not be restricted to externalizing behaviors such as opposition and aggression as it was also observed for shyness and social withdrawal (i.e., children with no preschool child-care experience manifested progressively lower levels of both behaviors). Furthermore, children who received child-care services during the preschool years, despite being more aggressive and oppositional in kindergarten, were not less prosocial than children who remained in parental care.

Although our data are consistent with both the diffusion hypothesis and the social group adaptation hypothesis for opposition and aggression, the two hypotheses differ conceptually and with respect to policy implications. According to the diffusion hypothesis, nonparental child-care services represent an early risk factor that increases the likelihood of later aggression so that, at a population level, lower prevalence of child care use implies lower prevalence of aggression. Policy implications would be able to reduce the prevalence of child-care use, particularly in the early years (see Jacob, 2009). However, according to the social group adaptation hypothesis, lower use of child-care services during the preschool years only postpones the social transition, as children go through the adaptation process later in child care or in kindergarten, as shown in the present study (see Côté et al., 2010, for a similar developmental phenomenon concerning infections). In other words, social group adaptation is accompanied by changes in social behaviors whether this transition happens earlier or later. As such, policy implications would rather be to facilitate the transition to social groups whenever it happens.

One approach to testing the diffusion and social group adaptation hypotheses would be to use experimental studies of kindergarten classrooms with different ratios of children with and without extensive center-based child-care experience (Linting & van IJzendoorn, 2009). Indeed, if a kindergarten classroom included only children with exclusive parental care during the preschool years, there could be no diffusion of aggression, and therefore, aggression levels should not change (e.g., Figure 2d, the exclusive parental care group would not have this sharp rise in aggression during the first 2 years of elementary school). Conversely, the social group adaptation hypothesis would still predict an initial increase in aggression levels in this scenario due to the adaptation to the social group setting. In addition, the social group adaptation hypothesis would predict that the levels of aggression of the newcomers depend on the levels of aggression of the preexisting group. If only a few children with previous extensive experience in child care join a classroom with a large number of children previously in exclusive parental care, a similar "regression to the group mean" should be observed. In such a case, the levels of aggression in children with previous extensive child-care experience could decrease, instead of remaining high and diffusing to other children.

According to this social group adaptation perspective, it is also possible to understand why some studies did not detect associations between child-care experience and externalizing behaviors (Linting & van IJzendoorn, 2009; Zachrisson, Dearing, Lekhal, & Toppelberg, 2013). In some environments, for instance, small groups of peers with a high adult-to-child ratio (McCartney et al., 2010), children may rapidly learn alternatives to aggression, and thus, the impact of child-care services on aggression may be small and/or of short duration; children joining later would adapt to a social group where aggression levels are already low.

Finally, the social group adaptation hypothesis does not exclude potential long-term effects of child care. Even if social transitions should impact social behaviors irrespective of age, it is still possible that timing matters. For instance, a very early transition may have long-term adverse or beneficial effects (e.g., children having entered child care at the 5 months assessment were still statistically significantly less shy at age 12 years although the size of the effect had decreased). However, the effects of child-care services should still partly fade away with time as all children undergo a social transition at some point.

Assessing Child-Care Effects: Further Perspectives

As shown in this study, short- and long-term developmental perspectives based on repeated assessments over time are essential in assessing child-care effects. Some of the "catch-up" effects reported in this study were noticed within only 2 years. A previous study even suggested that changes take place in the first months of child-care experience (Fein et al., 1993). As such, close repeated assessments in the 1st year of child-care experience may provide new perspectives on child-care effects. Second, the use of broad categories like internalizing and externalizing behavior may hide important effects. We propose to focus on behaviors that have a clear social dimension as they may be more affected by child care (e.g., focus on social anxiety instead of anxiety). Finally, more direct measures of the amount and nature (e.g., aggressive vs. nonaggressive) of peer interactions inside and outside child-care settings will help to test the alternative hypotheses discussed above.

Limitations

The long follow-up and use of teacher ratings yielded attrition. As such, the study sample should not be considered representative and attrition may have biased the results. We were unable to account for the role of child-care quality. The literature is mixed regarding the role played by child-care quality in the prediction of externalizing behaviors, with, overall, null or small effect sizes (Belsky, 2006; Belsky et al., 2007). However, quality may have played a role for other social behaviors. Finally, although the initial contributions of child-care variables to social behaviors had dissipated by age 12 years, "sleeper effects" may emerge at a later developmental stage (Belsky et al., 2007; Vandell, Belsky, Burchinal, Steinberg, & Vandergrift, 2010).

References

Achenbach, T., Edelbrock, C., & Howell, C. (1987). Empirically based assessment of the behavioral/emotional problems of 2- and 3-year-old children. *Journal of Abnormal Child Psychology*, *15*, 629–650. doi:10.1007/bf00917246

Asendorpf, J. B. (1993). Beyond temperament: A two-factorial coping model of the development of inhibition during childhood. In K. H. Rubin & J. B. Asendorpf (Eds.), *Social withdrawal, inhibition, and shyness in childhood* (pp. 265–289). Hillsdale, NJ: Erlbaum.

Averdijk, M., Besemer, S., Eisner, M., Bijleveld, C., & Ribeaud, D. (2011). The relationship between quantity, type, and timing of external childcare and child problem behaviour in Switzerland. *European Journal of Developmental Psychology*, *8*, 637–660. doi:10.1080/17405629.2011.571846

Aviezer, O., & Sagi-Schwartz, A. (2008). Attachment and non-maternal care: Towards contextualizing the quantity versus quality debate. *Attachment & Human Development*, *10*, 275–285. doi:10.1080/14616730802366699

Bates, J. E., Freeland, C. A., & Lounsbury, M. L. (1979). Measurement of infant difficultness. *Child Development*, *50*, 794–803. doi:10.2307/1128946

Belsky, J. (2006). Early child care and early child development: Major findings of the NICHD Study of Early Child Care. *European Journal of Developmental Psychology*, *3*, 95–110. doi:10.1080/17405620600557755

Belsky, J. (2009). Classroom composition, childcare history and social development: Are childcare effects disappearing or spreading? *Social Development*, *18*, 230–238. doi:10.1111/j.1467-9507.2008.00511.x

Belsky, J., Vandell, D. L., Burchinal, M., Clarke-Stewart, K. A., McCartney, K., & Owen, M. T., & the NICHD Early Child Care Research Network. (2007). Are there long-term effects of early child care? *Child Development*, *78*, 681–701. doi:10.1111/j.1467-8624.2007.01021.x

Boivin, M., Pérusse, D., Dionne, G., Saysset, V., Zoccolillo, M., Tarabulsy, G. M., . . . Tremblay, R. E. (2005). The genetic-environmental etiology of parents' perceptions and self-assessed behaviours toward their 5-month-old infants in a large twin and singleton sample. *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 46*, 612–630. doi:10.1111/j.1469-7610.2004.00375.x

Boivin, M., Vitaro, F., & Poulin, F. (2005). Peer relationships and the development of aggressive behavior in early childhood. In R. E. Tremblay, W. W. Hartup, & J. Archer (Eds.), *Developmental origins of aggression* (pp. 376–397). New York, NY: Guilford.

Caldwell, B. M., & Bradley, R. H. (1984). *Home Observation for Measurement of the Environment*. Little Rock: University of Arkansas.

Côté, S. M., Boivin, M., Nagin, D. S., Japel, C., Xu, Q., Zoccolillo, M., . . . Tremblay, R. E. (2007). The role of maternal education and nonmaternal care services in the prevention of children's physical aggression problems. *Archives of General Psychiatry, 64*, 1305–1312. doi:10.1001/archpsyc.64.11.1305

Côté, S. M., Petitclerc, A., Raynault, M.-F., Xu, Q., Falissard, B., Boivin, M., & Tremblay, R. E. (2010). Short- and long-term risk of infections as a function of group child care attendance: An 8-year population-based study. *Archives of Pediatrics & Adolescent Medicine, 164*, 1132–1137. doi:10.1001/archpediatrics.2010.216

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods, 1*(1), 16–29. doi:10.1037/1082-989x.1.1.16

Dallaire, D. H., & Weinraub, M. (2007). Infant-mother attachment security and children's anxiety and aggression at first grade. *Journal of Applied Developmental Psychology, 28*, 477–492. doi:10.1016/j.appdev.2007.06.005

Fabes, R. A., Hanish, L. D., & Martin, C. L. (2003). Children at play: The role of peers in understanding the effects of child care. *Child Development, 74*, 1039–1043. doi:10.1111/1467-8624.00586

Fein, G. G., Gariboldi, A., & Boni, R. (1993). The adjustment of infants and toddlers to group care: The first 6 months. *Early Childhood Research Quarterly, 8*, 1–14. doi:10.1016/S0885-2006(05)80095-X

Genolini, C., & Falissard, B. (2011). KmL: A package to cluster longitudinal data. *Computer Methods and Programs in Biomedicine, 104*, e112–e121. doi:10.1016/j.cmpb.2011.05.008

Geoffroy, M.-C., Séguin, J. R., Lacourse, E., Boivin, M., Tremblay, R. E., & Côté, S. M. (2012). Parental characteristics associated with childcare use during the first 4 years of life: Results from a representative cohort of Québec families. *Canadian Journal of Public Health, 103*, 76–80.

Goldstein, N. E., Arnold, D. H., Rosenberg, J. L., Stowe, R. M., & Ortiz, C. (2001). Contagion of aggression in day care classrooms as a function of peer and teacher responses. *Journal of Educational Psychology, 93*, 708–719. doi:10.1037/0022-0663.93.4.708

Guedeney, A., Foucault, C., Bougen, E., Larroque, B., & Mentré, F. (2008). Screening for risk factors of relational withdrawal behaviour in infants aged 14–18 months. *European Psychiatry, 23*, 150–155. doi:10.1016/j.eurpsy.2007.07.008

Hancock, G. R. (2001). Effect size, power, and sample size determination for structured means modeling and mimic approaches to between-groups hypothesis testing of means on a single latent construct. *Psychometrika, 66*, 373–388. doi:10.1007/BF02294440

Harlow, H. F. (1972). Love created, love destroyed, love regained. In *Modèles animaux du comportement humain [Animal models of human behaviour]* (pp. 13–60). Paris, France: Editions du CNRS.

Hay, D. F. (2005). The beginnings of aggression in infancy. In R. E. Tremblay, W. W. Hartup, & J. Archer (Eds.), *Developmental origins of aggression* (pp. 107–132). New York, NY: Guilford.

Hay, D. F., & Ross, H. S. (1982). The social nature of early conflict. *Child Development, 53*, 105–113. doi:10.2307/1129642

Heiser, N. A., Turner, S. M., & Beidel, D. C. (2003). Shyness: Relationship to social phobia and other psychiatric disorders. *Behaviour Research and Therapy, 41*, 209–221. doi:10.1016/s0005-7967(02)00003-7

Herba, C. M., Tremblay, R. E., Boivin, M., Liu, X., Mongeau, C., Séguin, J. R., & Côté, S. M. (2013). Maternal depressive symptoms and children's emotional problems: Can early child care help children of depressed mothers? *JAMA Psychiatry, 70*, 830–838. doi:10.1001/jamapsychiatry.2013.1361

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis, 15*, 199–236. doi:10.1093/pan/mpl013

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2011). MatchIt: Nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software, 42*, 1–28.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*, 1–55. doi:10.1080/10705519909540118

Imai, K., King, G., & Stuart, E. A. (2008). Misunderstandings between experimentalists and observationalists about causal inference. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 171*, 481–502. doi:10.1111/j.1467-985X.2007.00527.x

Jacob, J. I. (2009). The socio-emotional effects of non-maternal childcare on children in the USA: A critical review of recent studies. *Early Child Development and Care, 179*, 559–570. doi:10.1080/03004430701292988

Japel, C., Tremblay, R. E., & Côté, S. (2005). The quality of child care services: Results of the Québec Longitudinal Study of Child Development. *Éducation et Francophonie, II, I*(2), 7–27.

Jetté, M., & Des Groseilliers, L. (2000). *Survey description and methodology of the Longitudinal Study of Child Development in Québec (ELDEQ 1998–2002)*. Québec, Canada: Institut de la Statistique du Québec.

Kingery, J. N., Erdley, C. A., Marshall, K. C., Whitaker, K. G., & Reuter, T. R. (2010). Peer experiences of anxious and socially withdrawn youth: An integrative review of the developmental and clinical literature. *Clinical Child and Family Psychology Review, 13,* 91–128. doi:10.1007/s10567-009-0063-2

Kline, R. B. (2011). *Principles and practice of structural equation modeling*. New York, NY: Guilford.

Lamb, M. E., & Ahnert, L. (2006). Nonparental child care: Context, concepts, correlates, and consequences. In W. Damon, R. M. Lerner, K. A. Renninger, & I. E. Sigel (Eds.), *Handbook of child psychology: Vol. 4. Child psychology in practice,* 6th ed., pp. 950–1016). Hoboken, NJ: Wiley.

Liang, H., Pickles, A., Wood, N., & Simonoff, E. (2012). Adolescent emotional and behavioural outcomes of nonparental preschool childcare. *Social Psychiatry and Psychiatric Epidemiology, 47,* 399-407. doi:10.1007/s00127-011-0351-2

Linting, M., & van IJzendoorn, M. H. (2009). Dissipating or diffusing aggression after non-maternal childcare? Commentary on a novel hypothesis. *Social Development, 18,* 239–246. doi:10.1111/j.1467-9507.2008.00512.x

McCartney, K., Burchinal, M., Clarke-Stewart, A., Clarke-Stewart, A., Bub, K. L., Owen, M. T., . . . NICHD Early Child Care Research Network. (2010). Testing a series of causal propositions relating time in child care to children's externalizing behavior. *Developmental Psychology, 46,* 1–17. doi:10.1037/a0017886

Nærde, A., Ogden, T., Janson, H., & Zachrisson, H. D. (2014). Normative development of physical aggression from 8 to 26 months. *Developmental Psychology, 50,* 1710–1720. doi:10.1037/a0036324

National Institute of Child Health and Human Development Early Child Care Research Network. (2001). Child-care and family predictors of preschool attachment and stability from infancy. *Developmental Psychology, 37,* 847–862. doi:10.1037//0012-1649.37.6.847

National Institute of Child Health and Human Development Early Child Care Research Network. (2004). Trajectories of physical aggression from toddlerhood to middle childhood: Predictors, correlates, and outcomes. *Monographs of the Society for Research in Child Development, 69*(4). doi:10.1111/j.0037-976X.2004.00312.x

Oberski, D. (2014). lavaan.survey: An R Package for complex survey analysis of structural equation models. *Journal of Statistical Software, 57,* 1–27.

Offord, D., Boyle, M., & Racine, Y. (1989). Ontario Child Health Study: Correlates of disorder. *Journal of the American Academy of Child and Adolescent Psychiatry, 28,* 856–860.

Patterson, G. R., Reid, J. B., & Eddy, J. M. (2002). A brief history of the Oregon model. In J. B. Reid, G. R. Patterson, & J. J. Snyder (Eds.), *Antisocial behavior in children and adolescents. A developmental analysis and model for intervention* (pp. 3–21). Washington, DC: American Psychological Association.

Pineo, P. C., Porter, J., & Mcroberts, H. A. (1977). The 1971 Census and the socioeconomic classification of occupations*. *Canadian Review of Sociology/Revue Canadienne de Sociologie, 14,* 91–102. doi:10.1111/j.1755-618X.1977.tb00333.x

Pingault, J.-B., & Blatier, C. (2013). An observational study of the mother's role in regulating child conflict. *Devenir, 25,* 93–116.

Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1,* 385–401. doi:10.1177/014662167700100306

Rapee, R. M., Schniering, C. A., & Hudson, J. L. (2009). Anxiety disorders during childhood and adolescence: Origins and treatment. *Annual Review of Clinical Psychology, 5,* 311–341. doi:10.1146/annurev.clinpsy.032408.153628

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–36.

Rubin, K. H., Coplan, R. J., & Bowker, J. C. (2009). Social withdrawal in childhood. *Annual Review of Psychology, 60,* 141–171. doi:10.1146/annurev.psych.60.110707.163642

Sagi, A., Koren-Karie, N., Gini, M., Ziv, Y., & Joels, T. (2002). Shedding further light on the effects of various types and quality of early child care on infant-mother attachment relationship: The Haifa Study of Early Child Care. *Child Development, 73,* 1166–1186. doi:10.1111/1467-8624.00465

Séguin, L., Kantiébo, M., Xu, Q., Zunzunegui, M.-V., Potvin, L., Frohlich, K., & Dumas, C. (2001). Standard of living, health and development: Part I. Poverty, health conditions at birth and infant health. In *Longitudinal study of child development in Québec (ELDEQ 1998–2002)* (Vol. 1, No. 3, pp. 25–60). Québec, Canada: Institut de la statistique du Québec.

Statistics Canada. (1981). *Standard occupational classification 1980.* Ottawa, Canada: Supply and Services Canada.

Statistics Canada. (1995). *Overview of survey instruments for 1994–95 data collection, cycle 1.* Ottawa, Canada: Author.

Strayer, F. F., & Strayer, J. (1976). An ethological analysis of social agonism and dominance relations among preschool children. *Child Development, 47,* 980–989. doi:10.2307/1128434

Thibault, J., Jetté, M., & Desrosiers, H. (2001). Survey description and methodology. In *Longitudinal study of child development in Québec (ÉLDEQ 1998-2002)* (Vol. 1, No. 1). Québec, Canada: Institut de la statistique du Québec.

Tremblay, R. E. (2010). Developmental origins of disruptive behaviour problems: The "original sin" hypothesis, epigenetics and their consequences for prevention. *Journal of Child Psychology and Psychiatry, 51,* 341–367. doi:10.1111/j.1469-7610.2010.02211.x

Tremblay, R. E., Desmarais-Gervais, L., Gagnon, C., & Charlebois, P. (1987). The Preschool Behavior Questionnaire: Stability of its factor structure between cultures, sexes, ages and socioeconomic classes. *International Journal of Behavioral Development*, *10*, 467–484. doi:10.1177/016502548701000406

Van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, *45*, 1–67.

Vandell, D. L., Belsky, J., Burchinal, M., Steinberg, L., & Vandergrift, N. (2010). Do effects of early child care extend to age 15 years? Results from the NICHD study of early child care and youth development. *Child Development*, *81*, 737–756. doi:10.1111/j.1467-8624.2010.01431.x

Vitaro, F., Desmarais-Gervais, L., Tremblay, R. E., & Gag-non, C. (1992). Social adaptation at school entry and child care in preschool. *Revue Des Sciences de L'éducation*, *18*(1), 1–15.

Willms, J., & Shields, M. (1996). *A measure of socioeconomic status for the National Longitudinal Survey of Children*. Atlantic Center for Policy Research in Education.

Zachrisson, H. D., Dearing, E., Lekhal, R., & Toppelberg, C. O. (2013). Little evidence that time in child care causes externalizing problems during early childhood in Norway. *Child Development*, *84*, 1152–1170. doi:10.1111/cdev.12040

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's website:

Appendix S1. Early Nonparental Care and Social Behavior in Elementary School: Support for a Social Group Adaptation Hypothesis

Table S1. Variables Included in the Socioeconomic Status Composite

Table S2. Fit Indices for All Models

Table S3. Contribution Child Care Use and Covariates to Childhood Shyness

Table S4. Contribution of Child Care Use and Covariates to Childhood Social Withdrawal

Table S5. Contribution of Child Care Use and Covariates to Childhood Prosocial Behaviors

Table S6. Contribution of Child Care Use and Covariates to Childhood Opposition

Table S7. Contribution of Child Care Use and Covariates to Childhood Aggression

Table S8. Complementary Analyses Regarding the Type of Child Care

Table S9. Complementary Analyses With Multiple Child Care Variables

Table S10. Balance Between Using and Non-Using Children

Table S11. Contribution of Child Care Use Based on Propensity Score Matching

Table S12. Balance Between Low Trajectory and Rising/High Trajectories

Table S13. Contribution of Trajectories of Hours in Child Care Based on Propensity Score Matching

Table S14. Balance Between Individual and Center Care

Table S15. Contribution of Individual Child Care Versus Center Child Care Based on Propensity Score Matching

Child Development, September/October 2015, Volume 86, Number 5, Pages 1489–1506

Relationship Reciprocation Modulates Resource Allocation in Adolescent Social Networks: Developmental Effects

Stephanie Burnett Heyes
University of Oxford and University of Birmingham

Yeou-Rong Jih
University of Oxford

Per Block
University of Oxford and ETH Zürich

Chii-Fen Hiu
University of Oxford

Emily A. Holmes
MRC Cognition and Brain Sciences Unit
and Karolinska Institutet

Jennifer Y. F. Lau
University of Oxford and King's College London

Adolescence is characterized as a period of social reorientation toward peer relationships, entailing the emergence of sophisticated social abilities. Two studies (Study 1: $N = 42$, ages 13–17; Study 2: $N = 81$, ages 13–16) investigated age group differences in the impact of relationship reciprocation within school-based social networks on an experimental measure of cooperation behavior. Results suggest development between mid- and late adolescence in the extent to which reciprocation of social ties predicted resource allocation. With increasing age group, investment decisions increasingly reflected the degree to which peers reciprocated feelings of friendship. This result may reflect social-cognitive development, which could facilitate the ability to navigate an increasingly complex social world in adolescence and promote positive and enduring relationships into adulthood.

Adolescence, the period between childhood and adulthood, is characterized by changes in social behavior (Spear, 2000; Steinberg, 2008). Across several species, individuals in adolescence undergo a process of "social reorientation" entailing increased affiliation with peers, relative to family members (Lorme, Bell, & Sisk, 2013; Nelson, Leibenluft, McClure, & Pine, 2005).

During human adolescence, important changes take place within peer relationships. Increasing emphasis is placed on friendship intimacy and reciprocity (Youniss, 1982), and there is a rise in the importance of loyalty, commitment, and trust (Bigelow & La Gaipa, 1980). These changes in expectations of friendship are thought to impact the way in which adolescents manage and define their peer relationships. For example, an observed decrease in the number of self-reported close friends during adolescence is hypothesized to result from increased selectivity due to greater demands for intimacy and reciprocity (Urberg, Değirmencioğlu, Tolson, & Halliday-Scher, 1995). Taken together, this body of work suggests that during adolescence, changing perceptions and awareness of friendship qualities such as reciprocity may influence the way in which adolescents interact within their peer relationships.

It is proposed that the changes described above, which are observed in the context of adolescent

---

Stephanie Burnett Heyes, Yeou-Rong Jih, and Per Block each provided equal contribution.

Funding for this research came from a British Academy Postdoctoral Fellowship to Stephanie Burnett Heyes, a grant to Jennifer Y. F. Lau from the Calleva Research Centre for Evolution and Human Science, and research funding to Yeou-Rong Jih from the University of Oxford Masters in Psychological Research. Emily A. Holmes receives support from the Medical Research Council (United Kingdom) intramural programme (MC-A060-5PR50), a Wellcome Trust Clinical Fellowship (WT088217), and the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre based at Oxford University Hospitals NHS Trust, Oxford University. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, or the Department of Health. Thanks to Freya Harrison, Anneke Haddad, and Brian Parkinson for early discussions on the design of the study and to Charlotte Coyte and Alexandra McIntyre for help with data entry.

Correspondence concerning this article should be addressed to Stephanie Burnett Heyes, School of Psychology, University of Birmingham, Edgbaston, Birmingham, B15 2TT, United Kingdom. Electronic mail may be sent to burstephanie@gmail.com.

© 2015 The Authors
Child Development published by Wiley Periodicals, Inc. on behalf of Society for Research in Child Development.
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
All rights reserved. 0009-3920/2015/8605-0013
DOI: 10.1111/cdev.12396

peer relationships, are contingent upon developmental advances in social perspective taking (Selman, 1980). In one study (Selman, 1980, pp. 43–46) open-ended interview responses to social dilemmas were elicited from participants aged 9–17 years. Responses were coded with reference to a metric of the highest level of perspective taking used with competence, where Level 2 indicates simple reciprocal perspective taking (ability to grasp self *or* other's perspective), Level 3 indicates mutual perspective taking (ability to simultaneously grasp self *and* other's perspective), and Level 4 is indicative of a conceptual, multilevel understanding of interpersonal relations. From childhood to early adolescence (from ages 9–10 to 12–14 years), there was an increase in the proportion of Level 3 as compared to Level 2 responses, and transitioning into late adolescence (age 16–17 years), responses were predominantly at Levels 3 and 4. Thus, the study provided initial evidence that social perspective taking continue to develop during adolescence (see also Gurucharri & Selman, 1982).

Converging evidence from experimental paradigms suggests protracted development during adolescence in social-cognition (Burnett & Blakemore, 2009). In one study, child, adolescent, and adult participants (aged 7–27 years) played a computerized perspective-taking game in which their task was to move objects between shelves as instructed by a computer avatar. Results showed an increase throughout adolescence and into early adulthood in the tendency to automatically take into account the perspective or communicative intention of the computer avatar (Dumontheil, Apperly, & Blakemore, 2010). Other experimental studies have utilized game theory paradigms, in which principles such as fairness and altruism are operationalized via conversion into a common monetary currency that participants share, cooperate, or compete for during structured interactions (Rilling & Sanfey, 2011). Using these paradigms, studies have shown evidence for development during late childhood and adolescence in the tendency to take into account the perspectives, motives, and relative contributions of self and other (Almås, Cappelen, Sørensen, & Tungodden, 2010; Güroğlu, van den Bos, & Crone, 2009; Van den Bos, Westenberg, van Dijk, & Crone, 2010). For example, one study used a modified Ultimatum Game to show that investment decisions were increasingly modulated by the perceived ability of an unknown peer to reject selfish offers between ages 9 and 18 years (Güroğlu et al., 2009, Experiment 2). Results from these experimental studies align with the observational and structured interview findings summarized above suggesting advances in certain social-cognitive abilities up to and throughout the period of adolescence. However, while prior experimental studies provide important insights, they primarily rely upon interactions with unknown peers or computer stooges. Experimental research is needed that systematically investigates laboratory-based measures of social behavior (e.g., from game theory) in the context of real-world adolescent peer relationships.

A more nuanced empirical picture may emerge when social behavior is investigated in the context of authentic, established relationships. In authentic relationships, interactions unfold across extended periods of time, and individuals derive long-term benefits of sharing and cooperating when they can expect future direct or indirect reciprocation from others (Frank, 1998; Gardner & Foster, 2008; Hammerstein & Leimar, 2006; Nowak & Sigmund, 1998). Furthermore, during child and adolescent development, peer relationships provide a platform to practice and refine advanced social skills of cooperation, competition, and compromise (Youniss, 1982). Therefore, investigating social behavior in the context of authentic peer relationships may enable not only more ecologically valid assessments of social cognitive abilities, but may also show greater sensitivity to subtle changes within a developmental period. To accomplish this empirical goal, paradigms are needed that are capable of quantifying relevant features of real-world peer relationships, and such paradigms should be tied to experimental measures of social behavior.

A recently developed paradigm combines quantitative social network analysis techniques for mapping real-world relationships (Dekker, Krackhardt, & Snijders, 2007) with an experimental measure of prosocial cooperation behavior (Harrison, Sciberras, & James, 2011), where prosocial behavior is defined as voluntary behavior that benefits others (Eisenberg, Fabes, & Spinrad, 2007). In this recent study, adult members of a real-world social network (a research laboratory) completed an anonymous Social Network Questionnaire (SNQ) that comprised questions assessing feelings of closeness and companionship toward network peers. Responses were summed to yield a continuous measure of the perceived strength of social ties linking pairs of individuals (dyads) within the network, as well as the degree to which perceived social tie strength was reciprocated (Harrison et al., 2011). Subsequently, participants had the opportunity to earn money for themselves and a subset of their network peers in a physical effort-based task. Results showed that the strength of perceived social ties linking each participant to their network peers predicted the amount of money participants were willing to earn for their peers, at physical cost to themselves. Furthermore, the extent to which social ties were reciprocated predicted additional variance in investment. That is, participants' resource allocation decisions were sensitive to the degree to which feelings of closeness were reciprocated by their peers.

In the current study, we adapted this paradigm to investigate development during adolescence in the extent to which reciprocation of social ties within school-based social networks modulated prosocial resource allocation. Participants completed an SNQ, modified from Harrison et al. (2011), to take into account features of adolescent school-based relationships. To increase feasibility in a school setting, our measure of resource allocation was a modified one-shot Dictator Game. In this simple game theory paradigm, participants decide how much of an endowment of goods (e.g., money) to give to others, at cost to themselves. Thus, the task measures the extent to which participants are willing to share resources. Previous research has shown that Dictator Game resource allocation decreases with increasing social distance from peers, suggesting that this measure is sensitive to the strength of social ties in real-world social networks (Goeree, McConnell, Mitchell, Tromp, & Yariv, 2010).

We conducted two studies. In Study 1 ($N = 42$), we mapped two social networks: midadolescent ($M_{\text{age}} = 14$) and late adolescent ($M_{\text{age}} = 17$). In Study 2 ($N = 81$), we mapped four social networks: midadolescent ($M_{\text{age}} = 14$), later midadolescent ($M_{\text{age}} = 15$), and late adolescent (two networks; $M_{\text{age}} = 16$). In Study 1, each modified Dictator Game (mDG) point was worth real money, so participants had to decide how much of an opportunity cost they were willing to bear in order to share points. In Study 2, money was fictitious. Across both studies, we predicted that relationship strength would modulate resource allocation across age groups; that is, participants would invest more in recipients to whom they reported stronger social ties (Goeree et al., 2010; Harrison et al., 2011). Based on the evidence reviewed above (Dumontheil et al., 2010; Selman, 1980; Youniss, 1982), in Study 1 we predicted that the reciprocation of social ties would modulate resource allocation differentially across age groups; that is, in late adolescence, but not in midadolescence, individuals would invest more in individuals who reported stronger *reciprocal* social ties. In Study 2, we predicted a replication of Study 1 findings; in addition, the inclusion of three developmental time points enabled us to track hypothesized continuous (e.g., linear) change within the period of adolescence in the behavioral impact of relationship reciprocation. That is, we predicted that the extent to which reciprocation of social ties on the SNQ would modulate Dictator Game resource allocation would increase with increasing age group. Finally, both studies explored the impact of additional variables (Study 1 and Study 2: Machiavellian orientation [Christie & Geis, 1970], duration of acquaintance; Study 2: affluence [Currie et al., 2004], self-reported empathy [Davis, 1980], and number of best friends present in the current network) that could theoretically modulate prosocial behavior in the current study (Edele, Dziobek, & Keller, 2013; Goeree et al., 2010; Nettle, Colléony, & Cockerill, 2011; Spitzer, Fischbacher, Herrnberger, Grön, & Fehr, 2007) and that could, if they differed significantly between groups, give rise to confounding differences in comparisons across age groups.

Study 1

Method

Participants

We included 42 participants from a coeducational private school in a village in South East England based on cohort year group (grade level): Year 9 (midadolescent network): $n = 23$ (11 female), M (SD) age = 14.46 (.29) years, range = 13.88–14.79; Year 12 (late adolescent network): $n = 19$ (9 female), M (SD) age = 17.22 (.26) years, range = 16.91–17.97 (see Table 1). Participants were selected in consultation with a teacher such that individuals within each network spent class time together each week as a group. We obtained parent or guardian consent for all participants. Data were collected during May 2012. The study was approved by the University of Oxford research ethics committee and was carried out in accordance with the provisions of the World Medical Association Declaration of Helsinki.

Measures

Modified Dictator Game. Participants played an mDG to measure how much money (in the form of shopping vouchers) they were willing to give to other network members, at cost to themselves. The task is structurally equivalent to a one-shot Dictator

Table 1
Study 1 Descriptives for Each Year Group Network

	Year 9 Midadolescence			Year 12 Late adolescence			Group comparison
	M (*SD*)	Range	*N*	*M* (*SD*)	Range	*N*	$t(df), p$
Age (years)	14.46 (0.26)	13.88 to 14.79	23	17.22 (0.29)	16.91 to 17.97	19	
Kiddie Mach	51.74 (7.10)	41 to 70	23	55.47 (10.4)	42 to 85	19	$t(40) = 1.38, p = .176$
mDG: Points to self	13.88 (27.50)	0 to 100	23	28.36 (35.30)	0 to 100	19	$t(34)^{a} = 1.46, p = .154$

	Density *M* (SEM)	Range	*N*	Density *M* (SEM)	Range	*N*	$t(df), p$
mDG: Points to others	3.91 (.10)	0 to 15	506	3.98 (.21)	0 to 22.22	342	$t(846) = .28, p = .388$
SNQ relationship strength[b]	0.41 (.01)	−.13 to 1	506	0.33 (.01)	−.25 to 1	342	$t(846) = 4.36, p < .001$
SNQ relationship reciprocation[c]	0.15 (.01)	−.88 to .88	506	0.15 (.01)	-.63 to .63	342	$t(846) = 1.59, p = .056$
Duration of acquaintance (years)[d]	3.04 (.12)	1 to 15	506	2.6 (.11)	.67 to 17.32	342	$t(846) = 3.01, p = .001$

Note. The upper half of the table shows nonnetwork descriptives while the lower half of the table shows descriptives for network data. Network data were compared across groups using *t* tests based on bootstrapped mean density and standard error, with degrees of freedom based on *N* observations across groups. mDG = modified Dictator Game; SNQ = Social Network Questionnaire.
[a]Degrees of freedom reduced to take into account inequality of variances. [b]Relationship strength values represent perceived strength of unidirectional social ties. Values close to one reflect strong relationships; negative values or values close to zero reflect negative or neutral perceived relationships, respectively. [c]Reciprocation values represent the extent to which relationships are positively reciprocated. Positive values reflect greater in- than out-link strength. [d]For Year 9, maximum duration of acquaintance exceeds maximum age. Most likely, a participant rounded up to the nearest year when reporting duration of acquaintance.

Game (see Goeree et al., 2010) with multiple recipients. We provided participants with a table in which all members of the network were listed. Each participant had 100 points (1 point = £0.01) to distribute across the list. Instructions stated that there was no minimum point limit for each individual. Therefore, participants could choose to keep all the money by allotting 100 points to themselves, or to share points with network peers. We informed participants that the total amount of money they would receive would be determined by summing the points they allotted themselves and points other network members gave to them.

Social Network Questionnaire. We adapted a SNQ (Harrison et al., 2011) for adolescent participants. Participants rated network members on eight SNQ items. Five items provided binary data (yes = 1, no = 0): (a) whether the participant had a familial tie with each network member, (b) whether the participant had a current romantic relationship with each network member, (c) whether the participant spent time outside of school with each network member, and (d) whether the participant had ever chosen to academically work with each network member; (e) whether the participant would choose to confide personal bad news in each network member, and Three items provided ordinal or continuous data: (f) nature of relationship with each network member (4 = *best friend*, 3 = *good friend*, 2 = *friend*, 1 = *acquaintance*, 0 = *no relationship*, −1 = *negative relationship*), (g) how much the participant trusted each network member, relative to a random person from another school (1 = *more than*, 0 = *the same as*, −1 = *less than*), and (h) duration of acquaintance with each network member (years). The five relationship strength items (c, d, e, f, g) assessed various constructs (e.g., companionship, intimacy) that contribute to relationship strength (Bagwell & Schmidt, 2013). Additional items (a, b, h) were included to measure potential confounding effects of direct familial ties, romantic relationships, and duration of acquaintance. Since there were only three distant familial ties and two romantic relationships across the sample, these two variables were dropped from subsequent analyses. Participants also provided information regarding their gender.

Kiddie Mach. The Kiddie Mach (Nachamie, 1970) is a 20-item self-report questionnaire for children and adolescents derived from the adult Machiavellianism Scale–IV (Christie & Geis, 1970), which assesses lack of faith in human nature, manipulation, dishonesty, and distrust. The measure was included to investigate potential effects of Machiavellian orientation on mDG investment. Items include: "Most people are good and kind" and "Sometimes you have to hurt other people to get what you want." Agreement with Machiavellian items is indicated on a scale from 1 (*strongly* *disagree*) to 5 (*strongly agree*). Non-Machiavellian items are reverse scored such that high scores indicate disagreement and therefore Machiavellianism. The measure shows good internal reliability in adolescent samples (Sutton & Keogh, 2001). Previously, greater scores on this measure have been associated with reduced prosocial resource allocation on an mDG (Spitzer et al., 2007).

Experimental Procedure

Participants in each group completed the measures using pen and paper in a large testing room that afforded privacy for each individual. The measures took approximately 30 min to complete. After testing, all participants received a £5 shopping voucher for their participation in the mDG and a separate task unrelated to the present research.

Statistical Procedures

Below, we describe data processing procedures, followed by statistical analyses.

Data Processing

Modified Dictator Game. Investment in the mDG was used to create the following variables. (a) *Points allocated to self.* For each year group, we calculated the number of points allocated to self. (b) *Points allocated across the network.* For each year group, we created matrices in which the amount donated by a given participant to their classmates was listed in rows and the amount received by a given participant from their classmates was listed in columns.

Social Network Questionnaire. Responses on the SNQ were used to create the following variable matrices. (a) *Relationship strength.* For each year group, a relationship strength matrix representing the strength of social ties reported by a given participant toward other members of their network was created by summing responses to all five SNQ relationship strength items and, for ease of interpretation, dividing by eight the highest possible summed score. Thus, relationship strength matrix values (i.e., link weights between −.25 and 1) represent the strength of each social relationship: Negative values or values closer to 0 reflect more negative or neutral relationships, while values closer to 1 reflect stronger relationships. Matrix rows represent out-links, or how strong a relationship a given participant reports toward others. Matrix columns represent in-links, or how strong a relationship others report toward the participant. (b) *Reciprocation.* For each year group, a reciprocation matrix encapsulating the extent to which individuals reciprocated social ties was created by subtracting the relationship strength matrix for each year group from the respective transposed relationship strength matrix. As such, the reciprocation matrix is symmetrical with inverted sign; that is, for every positive entry there is a negative entry with the same absolute value on the other side of the diagonal of the matrix, and the sum of all elements of the matrix is necessarily zero. Reciprocation matrices make explicit the extent to which dyadic relationship strength is reciprocated: More negative values reflect relationships for which out-link exceeds in-link strength, while more positive values reflect relationships for which in-link exceeds out-link strength. (c) *Gender giving and gender receiving.* For each year group we created "gender giving" and "gender receiving" matrices by entering individual gender values (0 = female, 1 = male) across matrix rows and columns respectively. (d) *Duration of acquaintance.* For each year group we created a duration of acquaintance matrix using the SNQ duration of acquaintance item.

Kiddie Mach. Total Mach scores were calculated for each participant. In addition, for each year group we created "Mach giving" and "Mach receiving" matrices by entering individual Mach scores across matrix rows and columns, respectively.

Statistical Analyses

The data analysis strategy was as follows. First, *t* tests were conducted to investigate descriptive differences across networks. Second, multiple regressions were conducted to evaluate the impact of network and individual difference variables on mDG investment separately in each year group. Third, multigroup multiple regression was conducted to evaluate the impact of network and individual difference variables identified as of potential interest in Steps 1 and 2 on investment across the entire sample, and to pinpoint whether this impact varied as a function of year group, as a proxy for developmental age.

Descriptive differences across networks. T tests were conducted comparing the following variables across groups: mDG points allocated to self, mDG points allocated across the network, SNQ relationship strength, SNQ reciprocation, duration of acquaintance, and total Mach score. For comparison of network data (i.e., mDG points allocated across the network, SNQ relationship strength, SNQ reciprocation, and duration of acquaintance) a bootstrapping approach was used to compute standard errors based on mean network density, and subsequently, *t* tests were conducted comparing mean network density between groups (Snijders &

Borgatti, 1999). Note that for network data, degrees of freedom are equal to the number of tie variables across networks. Statistical significance was $p < .05$ two-tailed.

Multiple regression predicting resource allocation within networks. Multiple regression was conducted to evaluate the relationship between social network and individual difference variables and mDG point allocation in each year group. This entailed the use of statistical procedures capable of accounting for statistical dependence among observations, since didactic interaction data in social networks are based upon nonindependent observations: Measuring participant X's relationship with all individuals in the network requires repeat observations of X. Thus, observations of X are nonindependent. We utilized a statistical method capable of addressing nonindependent data: multiple regression with quadratic assignment procedure (MRQAP; Dekker et al., 2007; Hemelrijk, 1990; Krackhardt, 1988) executed in UCINET (version 6.415; Borgatti, Everett, & Freeman, 2002). MRQAP implements linear regression for matrix data. Regression coefficients can be interpreted as in ordinary least squares regression, although standard errors are calculated differently. Briefly, MRQAP regresses a dependent variable matrix upon explanatory variable matrices to obtain regression coefficients. Subsequently, row by column positions of the residuals resulting from the regression are randomly permuted (i.e., double Dekker semipartialing) to obtain regression coefficients of the permutations, which form the sampling distribution under the null hypothesis. The original regression coefficients are compared to this null distribution for hypothesis testing. We utilized UCINET's MRQAP with double Dekker semipartialing and 10,000 permutations to regress Year 9 and Year 12 mDG point allocation matrices against their respective year group explanatory variables: SNQ relationship strength, SNQ reciprocation, gender giving, gender receiving, duration of acquaintance, Mach giving, and Mach receiving. Statistical significance was $p < .05$, two-tailed for nondirectional and one-tailed for directional hypotheses. Directional hypotheses related to effects on point allocation of SNQ social tie strength (Harrison et al., 2011), gender (Balliet, Li, Macfarlan, & Van Vugt, 2011), Mach (Spitzer et al., 2007), and SNQ reciprocation (Dumontheil et al., 2010; Güroğlu et al., 2009; Selman, 1980). We dropped nonsignificant variables in a stepwise manner, starting with the variable for which the two-tailed p value was closest to 1.

Multiple regression predicting resource allocation across networks. Multigroup multiple regression (i.e., MG-MRQAP) was conducted to test directly for interactions between year group and the relation between mDG point allocation and explanatory variables described above. This first entailed the creation of full-sample matrices for each of the dependent and explanatory variables described above. These data have a multilevel structure, with participants ($N = 42$) nested in classes (Year 9, Year 12). We estimated regression parameters of interest using linear regression with pooled data from both classes and calculated standard errors by permuting each of the two matrices. Note that for the estimation of multiple classes simultaneously, a different, more efficient permutation method was chosen that tends to overestimate the size of the standard errors. To account for nesting of participants in classes, the procedure allowed permutation of rows and columns within classes only; in addition, the procedure allowed intercepts for the different classes to vary independently (i.e., a fixed effects approach by class). Thus, we could examine a network level variable (year group) for interactions with explanatory variables (e.g., reciprocation). Multiplying all explanatory matrix values by the appropriate network level variables enabled us to test whether the effect size of an explanatory variable depended on a network level variable. We coded the network level variable "year group" 0 and 1 for participants in Years 9 and 12, respectively, and ran the analysis in R (R Development Core Team, 2011). As for single-group MRQAP, statistical significance was $p < .05$, two-tailed for nondirectional and one-tailed for directional hypotheses. Directional hypotheses related to effects on point allocation of SNQ relationship strength, gender, Mach, and the interaction between age group and SNQ reciprocation. We dropped nonsignificant variables in a stepwise manner, starting with the variable for which the two-tailed p value was closest to 1. An exception to this rule was variables specifying main effects for which the interaction term was still in the model. In this case, exclusion of the main effect took place after exclusion of the interaction term. Predictors entered in the initial regression model were (a) variables that differed significantly between groups based on t-test results comparing descriptives across groups, (b) variables that had an impact on investment in either group based on single-group MRQAP results, and (c) terms included to test for interactions between all predictor variables and group (for a full list of predictor variables, see content and footnote of Table 3).

Results

Descriptive Differences Across Networks

Groups did not differ on points allocated to self, $t(34) = 1.46$, $p = .154$; points allocated to others, $t(846) = .29$, $p = .388$; SNQ reciprocation, $t(846) = 1.59$, $p = .056$; or total Mach score, $t(40) = 1.38$, $p = .176$ (see Table 1). There were group differences in duration of acquaintance, $t(846) = 3.01$, $p = .001$, and SNQ relationship strength, $t(846) = 4.36$, $p < .001$, such that Year 9 participants had been mutually acquainted for longer than Year 12 participants and reported greater relationship strength on average (see Table 1; Figures S1 and S2 in the online Supporting Information depict Year 9 and Year 12 networks based on relationship strength).

Multiple Regression Predicting Resource Allocation Within Networks

Year 9 model. The final Year 9 point allocation MRQAP model ($R^2_{adj} = .36$, $p < .001$) included SNQ relationship strength ($b = 4.84$, $p < .001$), gender giving ($b = -.80$, $p = .037$), and gender receiving ($b = .39$, $p = .006$; see Table 2). Ceteris paribus, participants in Year 9 gave more points to those with whom they reported stronger social ties, female participants gave more points to others in their network than did male participants, and male participants received more points than did female participants.

Year 12 model. The final Year 12 point allocation model ($R^2_{adj} = .24$, $p < .001$) included SNQ relationship strength ($b = 6.06$, $p < .001$), reciprocation ($b = 2.09$, $p = .032$), gender giving ($b = -1.47$, $p = .027$), and Kiddie Mach giving ($b = -.07$, $p = .032$; see Table 2). Ceteris paribus, participants in Year 12 gave more points to those toward whom they reported stronger social ties and to individuals who positively reciprocated social ties (i.e., those who reported a stronger relationship with the participant than the participant reported toward them). Female participants gave more points to others in their network than did male participants, and individuals who scored higher on Mach gave fewer points to others in their network than did those who scored lower on this measure.

Table 2
Study 1 Multiple Regression Models Predicting Resource Allocation Within Networks

Independent variable	Unstandardized regression coefficient	Standardized regression coefficient	p value
Year 9: Midadolescence MRQAP final model: $R^2_{adj} = .36$, $p < .001$			
Intercept	2.17	.00	< .001
Relationship strength	4.84	.56	< .001[a]
Gender giving	−0.80	−.18	.037[a]
Gender receiving	0.39	.08	.006
Year 12: Late adolescence MRQAP final model: $R^2_{adj} = .24$, $p < .001$			
Intercept	6.76	.00	< .001
Relationship strength	6.06	.40	< .001[a]
Reciprocation	2.09	.12	.032[a]
Gender giving	−1.47	−.19	.027[a]
Mach giving	−0.07	−.19	.032[a]

Note. Multiple regression with quadratic assignment procedure (MRQAP) was used to predict modified Dictator Game point allocation based on individual difference and network variables. Independent variables eliminated from the Year 9 model due to nonsignificance were Mach receiving ($b = -.00$, $p = .499$), reciprocation ($b = -.09$, $p = .432$), Mach giving ($b = .01$, $p = .402$), and duration of acquaintance ($b = -.03$, $p = .209$). Independent variables eliminated from the Year 12 model due to nonsignificance were gender receiving ($b = .27$, $p = .251$), Mach receiving ($b = .03$, $p = .111$), and duration of acquaintance ($b = -.15$, $p = .088$).
[a]Denotes a one-tailed p value; all other p values are two-tailed.

Multiple Regression Predicting Resource Allocation Across Networks

We regressed mDG point allocation across the entire sample ($N = 42$) on explanatory variable matrices using MG-MRQAP. In the final point allocation model ($R^2_{adj} = .26$, $p < .001$), significant predictor variables were SNQ relationship strength ($b = 4.82$, $p < .001$), SNQ Reciprocation × Year Group ($b = 2.64$, $p = .040$), and gender giving ($b = -1.18$, $p = .003$; Table 3). Across the entire sample, ceteris paribus, participants gave more to those toward whom they reported stronger social ties, male participants gave fewer points to others in their network than did female participants, and, relative to Year 9 participants, Year 12 participants gave more to individuals who positively reciprocated social ties (see Table 3).

Discussion

Results indicate that in both midadolescent ($M_{age} = 14$) and late adolescent ($M_{age} = 17$) networks, individuals invested more in peers toward whom they reported stronger social ties. This is consistent with a previous study in adults, in which cooperative investment in a physical effort-based task increased as a function of dyadic social tie strength (Harrison et al., 2011). The current finding is also consistent with a previous study in adoles-

Table 3

Study 1 Multiple Regression Models Predicting Resource Allocation Across Networks

Independent variable	Unstandardized regression coefficient	p value
Study 1 MG-MRQAP model: $R^2_{adj} = .255$, $p < .001$		
Year 9 intercept	2.58	< .001
Year 12 intercept (reference Year 9)	−0.23	.552
Relationship strength	4.82	< .001[a]
Reciprocation	−0.17	.820
Reciprocation × Year group	2.64	.040[a]
Gender giving	−1.18	.003[a]

Note. Multigroup multiple regression with quadratic assignment procedure (MG-MRQAP) was implemented to predict modified Dictator Game point allocation based on individual difference and network variables and terms specifying their respective interactions with year group (Year 9 and Year 12). Independent variables eliminated from the model were Gender Receiving × Year Group ($b = -.02$, $p = .493$), Gender Giving × Year Group ($b = -.41$, $p = .354$), Duration of Acquaintance × Year Group ($b = -1.17$, $p = .134$), duration of acquaintance ($b = -.08$, $p = .078$), gender receiving ($b = .29$, $p = .909$), Mach Giving × Group ($b = -.09$, $p = .071$), Mach giving ($b = -.04$, $p = .105$), and Relationship Strength × Year Group ($b = 1.83$, $p = .062$). Note that for the multigroup models in the current study, it is not appropriate to report standardized regression coefficients. In order to investigate interactions, the Year 12 intercept and reciprocation terms were retained. That is, the procedure for eliminating variables was to exclude the variable with the largest p value, unless this belonged to a main effect for which the interaction term was still in the model. In this case, exclusion of main effects took place after exclusion of interaction terms.

[a] Denotes a one-tailed p value; all other p values are two-tailed.

cents, which showed that cooperative investment decreased with increasing social distance (Goeree et al., 2010). Additionally, in the current study, we found that, relative to the midadolescent network, individuals in the late adolescent network invested more in peers who *reciprocated* strong ties. That is, late adolescents, but not midadolescents, were behaviorally sensitive to the degree to which others reciprocated their feelings of friendship, and adjusted their resource allocation strategy accordingly.

An intuitive way to interpret these results is as follows. In the older, late adolescent (Year 12) network, giving is maximal when I like you a lot *and* when you like me a lot (e.g., social tie strength = 1 in both directions). Giving is least when I dislike you a lot *and* you dislike me a lot (e.g., social tie strength = −0.25 in both directions). Between these extremes there are gradients—the less I like somebody, the less I give to them, but also, the less I am liked by them, the less I give to them. In contrast, in the younger, midadolescent (Year 9) network, giving is simply a function of how much I like you, and does not vary as a function of how much *you* like *me*.

This result is consistent with observations that the understanding of interpersonal reciprocity and mutuality continues to develop during adolescence (Selman, 1980), and with experimental evidence that social cognitive abilities, including perspective taking, continue to develop during adolescence (Dumontheil et al., 2010; Güroğlu et al., 2009). The current data constitute interesting preliminary evidence for differences between mid- and late adolescent networks in behavioral sensitivity to relationship reciprocation in the context of authentic, face-to-face peer relationships. However, to dissociate hypothesized continuous (e.g., linear) developmental advances in the behavioral impact of reciprocity from the effects of baseline, uncontrolled differences in individual characteristics between networks, comparison of more than two developmental time points is needed. Therefore, Study 2 examined developmental reciprocity in greater detail by directly comparing three cross-sectional time points within a larger mid- to late adolescent sample. We predicted that the degree to which participants reciprocated social ties would modulate resource allocation to a greater extent in late adolescent, relative to midadolescent, networks.

In Study 1, two descriptive differences were observed across groups: First, duration of acquaintance was higher in midadolescence than in late adolescence. However, since this variable was eliminated due to nonsignificance from the multigroup regression model, it cannot account for group differences in patterns of investment. Second, SNQ relationship strength was greater in midadolescence than in late adolescence. However, since the interaction between age group and relationship strength was nonsignificant in the final multigroup regression model, the group difference in this variable cannot account for the pattern of findings. In Study 2, we explored the generalizability of this finding.

In Study 1, individual difference variables exerted subtle effects on investment patterns. Based on single- and multigroup regression models, females tended to give more points than males, consistent with evidence that females tend to be more cooperative than males in mixed-gender contexts (Balliet et al., 2011). Machiavellianism had a weak, negative impact on investment decisions in the late adolescent network, aligning with evidence that individuals scoring more highly on this measure value personal gain over benefits for others (Christie & Geis, 1970; Wilson, Near, & Miller,

1996) and with a previous study in adults using an mDG (Spitzer et al., 2007). However, the absence of an effect in the midadolescent network suggests that the current finding should be interpreted with caution. Furthermore, in addition to Machiavellianism, there are other individual characteristics that may influence investment decisions, such as wealth (Nettle et al., 2011), empathy (Edele et al., 2013), and the number of close friends present in the current network (Goeree et al., 2010), and these could also contribute to the observed differences in investment patterns between networks. Therefore, in Study 2 we included three additional measures: the Interpersonal Reactivity Index (IRI; Davis, 1980), the Family Affluence Scale (Currie et al., 2004), and a peer nominations questionnaire to investigate potential group differences in self-reported empathy, individual levels of wealth, and the number of best friends present in the current network, respectively, while a single-gender sample was used to increase statistical power by eliminating gender effects as observed in Study 1.

Study 2

Method

Participants

We recruited 81 participants from a single-sex (female) state school in a town in South East England based on cohort year group (grade level): Year 9: $n = 13$, M (SD) age = 14.10 (.36) years, range = 13.48–14.47; Year 10: $n = 24$, M (SD) age = 15.10 (.38) years, range = 14.53–16.09; Year 11A: $n = 25$, M (SD) age = 15.96 (.24) years, range = 15.50–16.38; Year 11B: $n = 19$, M (SD) age = 16.00 (.28) years, range = 15.58–16.41 (i.e., Year 11A and Year 11B networks are different classes from the same grade level). Participants were selected in consultation with a teacher such that individuals within each network spent class time together each week as a group. Data were collected during February 2013. We obtained parent or guardian consent for all participants. The study was approved by the University of Oxford research ethics committee and was carried out in accordance with the provisions of the World Medical Association Declaration of Helsinki.

Measures

Modified Dictator Game. Participants played an mDG to measure how much money they would be willing to give to other network members, at cost to themselves. The mDG was identical to that in Study 1 except we instructed participants to imagine they had £100 to distribute among their network. The use of hypothetical rather than real money was motivated by the preferences of the school from which the sample was recruited and based on evidence that hypothetical money effectively motivates participant behavior (Miyapuram, Tobler, Gregorios-Pippas, & Schultz, 2012).

Social Network Questionnaire. The SNQ was identical to that in Study 1, with the following three exceptions: (a) The duration of acquaintance item was altered for speed and ease of completion. Participants rated whether they had known each network member since primary school (= 1) or secondary school (= 0). This variable was labeled "Duration 1." Participants also reported for how long (in years) they had attended the current school. This variable was labeled "Duration 2." (b) The romantic ties item was dropped due to low positive response rates in Study 1. (c) An item was included to assess conflict resolution. Along with intimacy and companionship, conflict resolution is an important dimension of friendship and contributor to relationship strength (Bagwell & Schmidt, 2013). For each network member, participants rated how much effort they would expend to resolve a disagreement (1 = *great effort*, 0 = *small effort*, −1 = *would not care*). There were no familial ties across the sample, and therefore the familial ties variable was dropped from analysis.

Interpersonal Reactivity Index. The IRI (Davis, 1980) was administered to investigate the relationship between prosocial investment and self-reported empathy. The IRI is a 28-item multidimensional self-report measure of dispositional empathy with four 7-item subscales: perspective taking (PT; tendency to spontaneously adopt the psychological point of view of others in everyday life; e.g., "I sometimes try to understand my friends better by imagining how things look from their perspective"), empathic concern (EC; tendency to experience feelings of sympathy and compassion for unfortunate others; e.g., "I often have tender, concerned feelings for people less fortunate than me"), personal distress (PD; tendency to experience distress and discomfort in response to extreme distress in others; e.g., "Being in a tense emotional situation scares me"), and fantasy (F; tendency to imaginatively transpose oneself into fictional situations; e.g., "When I am reading an interesting story or novel, I imagine how I would feel if the events in the story were happening to me"). Statements are scored on a 5-point scale from A (*does not describe me very well*) to E (*describes me very well*). Good internal reliability and test–retest stability for each subscale have been reported for this measure in adults (Davis, 1980). Developmental research has shown a similar factor structure for this measure in adolescent participants, with age-associated increases in test–retest stability (Davis & Franzoi, 1991). Previous research has shown a relation between IRI EC and prosocial resource allocation (Edele et al., 2013). In the current study, we focused our analysis on the EC and PT subscales.

Family Affluence Scale II (FASII). The FASII (Currie et al., 2004) is a four-item measure of socioeconomic status suitable for completion by adolescents. This measure was administered to assess potential effects on investment of differing levels of wealth. Items include: "Does your family own a car or van?" (responses: No; Yes, one; Yes, two or more) and "Do you have your own bedroom?" (responses: No; Yes). Reliability and validity has been reported in detail elsewhere; for example, studies have shown high child–parent agreement (Currie et al., 2008). Previous research has shown an impact of differing levels of wealth on Dictator Game investment (Nettle et al., 2011).

Kiddie Mach. The Kiddie Mach (Nachamie, 1970) was administered as in Study 1.

Best Friends. A peer nominations questionnaire was administered in which participants listed their five best friends (in any setting). This measure was administered to assess potential effects on mDG investment of the number of self-reported close friendships in the current network.

Experimental Procedure

Participants in each group completed the measures using pen and paper in a large testing room that afforded privacy for each individual. The measures took approximately 40 min to complete.

Statistical Procedures

Below, we describe data processing procedures, followed by statistical analyses.

Data Processing

Modified Dictator Game. As in Study 1, we used mDG responses to create two variables: money allocation to self and money allocation to others. The latter is a matrix.

Social Network Questionnaire. Responses on the SNQ were used to create the following variables. (a) *Relationship strength.* For each year group, a relationship strength matrix representing the strength of social ties reported by a given participant toward other members of their network was created by summing responses to all six SNQ relationship strength items and, for ease of interpretation, dividing by nine the highest possible summed score. Thus, relationship strength matrix values (i.e., link weights between −.33 and 1) represent the strength of each social relationship. (b) *Reciprocation.* For each year group, a reciprocation matrix was created by subtracting the relationship strength matrix (out-link values) for each year group from a transposed relationship strength matrix (in-link values), as in Study 1. (c) *Duration of acquaintance 1.* SNQ Duration 1 responses (see Measures) were entered into a matrix labeled "Duration 1." (d) *Duration of acquaintance 2.* SNQ Duration 2 responses (see Measures) were entered into "Duration 2 giving" and "Duration 2 receiving" matrices.

Individual difference variables. We calculated total scores on the FASII and Kiddie Mach, and summed the number of best friends present in the current network. For the IRI, we calculated total scores on the EC and PT subscales. Subsequently, for each year group, we created separate "giving" and "receiving" matrices for each of these five measures by entering scores across matrix rows and columns respectively.

Statistical Analysis

The data analysis strategy was as in Study 1. First, statistical tests were conducted to investigate descriptive differences across networks. Second, multiple regressions were conducted to evaluate the impact of network and individual difference variables on mDG investment separately in each network. Third, multigroup multiple regression was conducted to evaluate the impact of network and individual difference variables identified as of potential interest in Steps 1 and 2 on investment across the entire sample, and to pinpoint whether this impact varied as a function of year group, as a proxy for developmental age.

Descriptive differences across networks. Analyses of variance (ANOVAs) were conducted comparing mDG money allocated to self, Duration 2, Mach, FASII, IRI PT, and IRI EC across groups, and a chi-square test was conducted comparing best friends across groups, with statistical significance set at $p < .05$ two-tailed. For comparison of network data (i.e., mDG money allocated to others, SNQ relationship strength, SNQ reciprocation and Duration 1) across groups, ANOVA is not permitted since there is no accepted method for computing whole sample and individual network variance based on nonindependent observations. Therefore, we used the bootstrapping method (see Study 1) to conduct t tests comparing network data between all *pairs* of networks (including Year 11A and Year 11B), with Bonferroni correction for multiple comparisons (statistical significance = .05/6, i.e., $p < .0008$ two-tailed).

Multiple regression predicting resource allocation within networks. As in Study 1, we used UCINET's MRQAP with double Dekker semipartialing and 10,000 permutations to regress Year 9, Year 10, Year 11A, and Year 11B mDG investment matrices against their respective year group explanatory variables: SNQ relationship strength, SNQ reciprocation, Duration 1, Mach giving, Mach receiving, Duration 2 giving, Duration 2 receiving, FASII giving, FASII receiving, IRI EC giving, IRI EC receiving, IRI PT giving, IRI PT receiving, best friends giving, and best friends receiving. Statistical significance and the strategy for eliminating nonsignificant variables were as in Study 1. Directional hypotheses were as per Study 1 single-group MRQAP with the addition of the effect on point allocation of the two IRI subscales (Edele et al., 2013).

Multiple regression predicting resource allocation across networks. MGMRQAP tested for interactions between school year group (as a proxy for developmental age) and relations between potential explanatory variables and investment in the mDG. As in Study 1, this first entailed the creation of full-sample matrices for each of the dependent and explanatory variables described above containing data from all participants ($N = 81$) across the four classes: Years 9, 10, 11A, and 11B. We coded the network level variable "year group" 0, 1, and 2 for participants in Years 9, 10, and 11, respectively. Statistical significance and the strategy for eliminating nonsignificant variables were as in Study 1. Directional hypotheses were as per Study 1 MG-MRQAP with the addition of the effect on point allocation of IRI subscales. Predictors entered in the initial regression model were (a) variables that differed significantly between groups based on initial ANOVA, chi-square, or t-test results comparing descriptives between groups; (b) variables that had an impact on investment in any group based on single-group MRQAP; and (c) terms included to test for interactions with group (for a full list of predictor variables, see content and footnote of Table 6).

Results

Descriptive Differences Across Networks

The four groups did not differ on fictitious money allocated to self, $F(3, 80) = .96$, $p = .415$; Mach, $F(3, 70) = .12$, $p = .949$; FASII, $F(3, 78) = 1.57$, $p = .203$; IRI PT, $F(3, 78) = .19$, $p = .905$; or IRI EC, $F(3, 78) = .04$, $p = .987$ (see Table 4). Groups differed on the number of years individuals had attended the current school: Duration 2, $F(3, 79) = 33.82$, $p < .001$, and the number of self-reported best friends present in the current network, $\chi^2 = 22.64$, $p = .007$. T tests comparing network data between pairs of groups showed differences in money allocated to others (Year 9 vs. all other groups: all three comparisons, $ps < .0008$), SNQ relationship strength (Year 11A vs. Year 9 and Year 11B: both $p < .0008$), SNQ reciprocation (Year 9 vs. Year 10 and Year 11A: both $ps < .0008$), and Duration 1 (Year 10 vs. Year 11B: $p < .0008$; for further details, see Table 4 and Appendix S1). Supplementary Figures S3–S6 depict Years 9, 10, 11A, and 11B networks based on relationship strength.

Multiple Regression Predicting Resource Allocation Within Networks

We regressed each year group's mDG allocation matrix on their explanatory variable matrices using MRQAP yielding final models shown in Table 5. For Year 9, $R^2_{adj} = .47$, $p < .001$; for Year 10, $R^2_{adj} = .39$, $p < .001$; for Year 11A, $R^2_{adj} = .39$, $p < .001$; and for Year 11B, $R^2_{adj} = .39$, $p < .001$. SNQ relationship strength was significant in all networks (Year 9: $b = 10.37$, $p < .001$; Year 10: $b = 12.16$, $p < .001$; Year 11A: $b = 15.87$, $p < .001$; Year 11B: $b = 19.84$, $p < .001$). SNQ reciprocation was significant in Year 10, Year 11A, and Year 11B networks only (Year 10: $b = 3.03$, $p = .002$; Year 11A: $b = 3.72$, $p = .002$; Year 11B: $b = 6.52$, $p < .001$). The impact of IRI EC on point allocation was significant in Year 9 only ($b = .23$, $p = .023$), and the impact of Duration 1 was significant in Year 10 only ($b = -1.64$, $p = .010$). Ceteris paribus, participants in all networks gave more points to those towards whom they reported stronger social ties. In Year 10 and both Year 11 networks, ceteris paribus, participants gave more to individuals who positively reciprocated social ties. In Year 9, ceteris paribus, participants who scored higher on the IRI EC subscale gave more to their classmates, and in Year 10, participants gave less to individuals they had known since primary versus secondary school (see Table 5).

Table 4
Study 2 Descriptives for Each Year Group Network

	Year 9 Midadolescence			Year 10 Later midadolescence			Year 11A Late adolescence			Year 11B Late adolescence			Group comparison
	M (*SD*)	Range	*N*	*M* (*SD*)	Range	*N*	*M* (*SD*)	Range	*N*	*M* (*SD*)	Range	*N*	*F*(*df*), *p* or χ^2_{df}, *p*
Age (years)	14.10 (0.36)	13.5 to 14.5	13	15.10 (0.38)	14.5 to 16.1	24	15.96 (0.24)	15.5 to 16.4	25	16.00 (0.28)	15.6 to 16.4	19	
Kiddie Mach	56.35 (5.08)	50 to 68	13	57.94 (10.82)	42 to 79	16	57.16 (7.61)	44 to 74	25	56.67 (6.42)	45 to 67	19	*F*(3, 70) = .12, *p* = .949
IRI PT	21.61 (3.87)	12 to 26	13	22.37 (5.99)	11 to 31	22	22.80 (4.43)	10 to 29	25	22.82 (5.72)	14 to 35	19	*F*(3, 78) = .19, *p* = .905
IRI EC	25.50 (3.82)	18 to 30	13	25.20 (5.56)	11 to 33	22	25.00 (4.82)	14 to 31	25	24.95 (3.75)	16 to 31	19	*F*(3, 78) = .04, *p* = .987
FASII	6.54 (2.37)	2 to 9	13	5.87 (2.01)	0 to 9	22	6.80 (1.56)	4 to 9	25	7.00 (1.50)	4 to 9	18	*F*(3, 78) = 1.57, *p* = .203
Duration 2	2.50 (0.00)	2.50 to 2.50	13	3.25 (0.74)	0.5 to 3.5	24	4.46 (0.20)	3.5 to 4.5	25	4.34 (0.69)	1.5 to 4.5	19	*F*(3, 80) = 53.20, *p* < .001
Best friends	0.91 (0.70)	0 to 2	12	0.96 (0.93)	0 to 3	23	0.80 (0.96)	0 to 3	25	1.00 (1.05)	0 to 3	19	$\chi^2(9)$ = 22.64, *p* = .007
mDG: Points to self	20.29 (18.98)	0 to 58.82	13	33.08 (32.33)	0 to 100	24	36.79 (30.24)	0 to 90	25	31.29 (27.53)	0 to 100	19	*F*(3, 80) = .96, *p* = .415

	Density *M* (SEM)	Range	*N*	Density *M* (SEM)	Range	*N*	Density *M* (SEM)	Range	*N*	Density *M* (SEM)	Range	*N*	*t*-test group difference [a]
mDG: Points to others	6.64 (.42)	0 to 30	156	2.78 (.24)	0 to 50	552	2.63 (.28)	0 to 60	600	3.82 (.46)	0 to 100	342	Year 9 versus all other groups
Relationship strength	0.22 (.03)	−.25 to 1	156	0.15 (.01)	−.33 to 1	552	0.10 (.01)	−.33 to 1	600	0.24 (.02)	−.33 to 1	342	Year 9 versus Year 11A and Year 11B
Relationship reciprocation	0.21 (.01)	−.81 to .81	156	0.15 (.01)	−.67 to .67	552	0.14 (.01)	−.78 to .78	600	0.16 (.01)	−.63 to .63	342	Year 9 versus Year 10 and Year 11A
Duration 1	0.12 (.03)	0 to 1	156	0.15 (.02)	0 to 1	552	0.10 (.01)	0 to 1	600	0.06 (.01)	0 to 1	342	Year 10 versus Year 11B

Note. The upper half of the table shows nonnetwork descriptives while the lower half of the table shows descriptives for network data. Network data were compared across groups using *t* tests based on bootstrapped mean density and standard error, with degrees of freedom based on *N* observations across groups. IRI PT/EC = interpersonal reactivity index PT/EC subscales; FASII = Family Affluence Scale; Duration 2 = for how many years the participant had attended the current school; Best friends = number of self-nominated best friends in the current network; mDG = modified Dictator Game; Duration 1 = whether the participant had known his or her classmate since primary (= 1) or secondary (= 0) school.

[a] For comparison of network data across groups, analyses of variance is not permitted since there is no accepted method for computing whole sample and individual network variance based on nonindependent observations. Therefore, we used the bootstrapping method (see Methods in Study 1) to conduct *t* tests comparing network data between all pairs of networks, with Bonferroni correction for multiple comparisons (statistical significance = .05/6, i.e., $p < .0008$ two-tailed). For *t* and *p* values, see Appendix S1.

Table 5

Study 2 Multiple Regression Models Predicting Resource Allocation Within Networks

Independent variable	Unstandardized regression coefficient	Standardized regression coefficient	p value
Year 9: Midadolescence			
MRQAP final model: $R^2_{\text{adj}} = .47$, $p < .001$			
Intercept	0.74	.00	< .001
Relationship strength	10.37	.66	< .001[a]
Empathic concern giving	0.23	.16	.023[a]
Year 10: Later midadolescence			
MRQAP final model: $R^2_{\text{adj}} = .39$, $p < .001$			
Intercept	1.23	.00	< .001
Relationship strength	12.16	.67	< .001[a]
Reciprocation	3.03	.11	.002[a]
Duration 1	−1.64	−.11	.010
Year 11A: Late adolescence			
MRQAP final model: $R^2_{\text{adj}} = .39$, $p < .001$			
Intercept	1.14	.00	< .001
Relationship strength	15.87	.66	< .001[a]
Reciprocation	3.72	.11	.002[a]
Year 11B: Late adolescence			
MRQAP final model: $R^2_{\text{adj}} = .39$, $p < .001$			
Intercept	−0.72	.00	< .001
Relationship strength	19.84	.68	< .001[a]
Reciprocation	6.52	.16	< .001[a]

Note. Multiple regression with quadratic assignment procedure (MRQAP) was used to predict modified Dictator Game point allocation based on individual difference and network variables. Independent variables eliminated from the Year 9 model due to nonsignificance were IRI EC receiving ($b = .01$, $p = .465$), IRI PT receiving ($b = .01$, $p = .471$), best friends receiving ($b = .07$, $p = .438$), Mach giving ($b = -.02$, $p = .429$), best friends giving ($b = -.27$, $p = .339$), Duration 1 ($b = -.40$, $p = .336$), IRI PT giving ($b = -.06$, $p = .275$), FASII receiving ($b = -.08$, $p = .262$), SNQ reciprocation ($b = .90$, $p = .252$), Mach receiving ($b = -.09$, $p = .082$), and FASII giving ($b = -.33$, $p = .054$). Due to low variance, Duration 2 variables were not entered into the model. Independent variables eliminated from the Year 10 model were Mach giving ($b = .00$, $p = .485$), IRI EC receiving ($b = .00$, $p = .468$), best friends giving ($b = .05$, $p = .470$), FASII giving ($b = .02$, $p = .445$), IRI PT giving ($b = -.09$, $p = .131$), IRI EC giving ($b = .04$, $p = .207$), Duration 2 receiving ($b = .29$, $p = .121$), best friends receiving ($b = -.19$, $p = .150$), Duration 2 giving ($b = -.65$, $p = .091$), IRI PT receiving ($b = .02$, $p = .074$), FASII receiving ($b = -.12$, $p = .059$), and Mach receiving ($b = .01$, $p = .091$). Independent variables eliminated from the Year 11A model were IRI EC receiving ($b = -.01$, $p = .446$), IRI EC giving ($b = -.02$, $p = .437$), FASII giving ($b = .06$, $p = .403$), best friends giving ($b = .06$, $p = .424$), FASII receiving ($b = -.10$, $p = .307$), best friends receiving ($b = -.13$, $p = .313$), IRI PT giving ($b = -.08$, $p = .183$), Mach giving ($b = -.02$, $p = .328$), IRI PT receiving ($b = -.11$, $p = .063$), Mach receiving ($b = -.03$, $p = .182$), and Duration 1 ($b = -1.26$, $p = .072$). Due to low variance, Duration 2 variables were not entered into the model. Independent variables eliminated from the Year 11B model were FASII receiving ($b = -.02$, $p = .461$), IRI EC giving ($b = .02$, $p = .457$), Mach giving ($b = .01$, $p = .405$), FASII giving ($b = -.06$, $p = .382$), IRI PT giving ($b = -.04$, $p = .310$), Mach receiving ($b = -.01$, $p = .317$), IRI EC receiving ($b = .05$, $p = .276$), best friends receiving ($b = -.18$, $p = .261$), best friends giving ($b = -.39$, $p = .156$), IRI PT receiving ($b = -.07$, $p = .076$), and Duration 1 ($b = -2.18$, $p = .090$). Due to low variance, Duration 2 variables were not entered into the model. SNQ = Social Network Questionnaire; IRI PT = interpersonal reactivity index perspective-taking; EC = empathic concern; FASII = family Affluence Scale.

[a] Denotes a one-tailed p value; all other p values are two-tailed.

Multiple Regression Predicting Resource Allocation Across Networks

We regressed mDG allocation across the entire sample ($N = 81$) on explanatory variable matrices using MG-MRQAP. In the final point allocation model ($R^2_{\text{adj}} = .43$, $p < .001$), significant predictor variables were SNQ relationship strength ($b = 9.45$, $p < .001$), SNQ Relationship Strength × Year Group ($b = 3.88$, $p < .001$), SNQ Reciprocation × Year Group ($b = 2.15$, $p = .010$), Duration 1 ($b = -1.32$, $p = .014$), IRI EC giving ($b = .22$, $p = .006$), and IRI EC Giving × Year Group ($b = -.13$, $p = .026$; see Table 6). Ceteris paribus, participants gave more fictitious money to those to whom they reported stronger social ties, with this effect being greater in older year groups. Relative to younger year groups, ceteris paribus, participants in older year groups gave more to individuals who positively reciprocated social ties. All else being equal, participants who scored higher on IRI EC gave more, with this effect greater in younger networks, while participants gave less to individuals they had known since primary versus secondary school.

Table 6
Study 2 Multiple Regression Models Predicting Resource Allocation Across Networks

Independent variable	Unstandardized regression coefficient	p value
MG-MRQAP model		
$R^2_{\text{adj}} = .43, p < .001$		
Year 9 intercept	0.53	.006
Year 10 intercept (reference Year 9)	−1.52	.126
Year 11A intercept (reference Year 9)	−1.88	.480
Year 11B intercept (reference Year 9)	1.01	< .001
Relationship strength	9.45	< .001[a]
Relationship Strength × Year Group	3.88	< .001[a]
Reciprocation	0.62	.658
Reciprocation × Year Group	2.15	.010[a]
Duration 1	−1.32	.014
Empathic concern giving	0.22	.006[a]
Empathic Concern Giving × Year Group	−0.13	.026

Note. Multigroup multiple regression with quadratic assignment procedure (MG-MRQAP) was implemented to predict modified Dictator Game point allocation based on individual difference and network variables and terms specifying their respective interactions with year group (Year 9, Year 10, and Year 11). Independent variables eliminated due to nonsignificance were Best Friends Giving × Group ($b = .04$, $p = .552$), best friends giving ($b = .00$, $p = .505$), Best Friends Receiving × Group ($b = .05$, $p = .554$), Duration 1 × Group ($b = -.26$, $p = .396$), Duration 2 Giving × Group ($b = .26$, $p = .661$), best friends receiving ($b = -.08$, $p = .316$), Duration 2 Receiving × Group ($b = -.61$, $p = .112$), Duration 2 receiving ($b = -.03$, $p = .440$), and Duration 2 giving ($b = -.46$, $p = .071$). Note that the nonsignificant variable reciprocation is retained to examine its interaction with year group.

[a] Denotes a one-tailed p value; all other p values are two-tailed.

Discussion

Study 2 was conducted to replicate findings from Study 1, and to extend them by investigating hypothesized development between mid- to late adolescence in the impact of reciprocation on mDG resource allocation. Thus, Study 2 compared three developmental time points to evaluate evidence for an effect of school year group, as a proxy for developmental age, on the relation between the social network variable reciprocation and mDG investment. Replicating Study 1 results, we found first that relationship strength modulated investment in all networks, whereas reciprocation modulated investment only in older networks (Years 10 and 11). Second, we found that with increasing age group, reciprocation exerted an *increasing* impact on investment. It is notable that such findings were obtained while taking into account potentially confounding variables such as duration of acquaintance, the number of best friends present in the current network, and familial levels of wealth.

In Study 2, we observed an interaction between year group and SNQ relationship strength that was not present in Study 1. That is, with increasing year group, resource allocation was increasingly modulated by the social tie strength reported by a participant toward his or her classmates, such that participants in older year groups were increasingly likely to give more to those to whom they reported a stronger social tie. Further studies are needed to replicate and confirm this novel age group difference in the impact of social tie strength on investment.

Some descriptive differences were observed across groups in Study 2. Duration 2, that is, for how many years the participant had attended the current school, was greater in older year groups (in contrast to greater mean duration of acquaintance in the younger vs. older group in Study 1). Additionally, the number of self-reported best friends present in the current network differed significantly across groups. However, since these variables were eliminated from the full sample regression model, they cannot be said to account for age group differences in patterns of investment. Relationship strength and reciprocation differed across groups (see Table 4 and Appendix S1), but not in a manner that was systematically related to age group. Fictitious money invested in others differed across groups; however, this difference is taken into account statistically within the multigroup model, since in a fixed effects approach by groups the intercept (amount invested in other network members if all independent variables are equal to 0) is allowed to vary between groups.

In addition to the above-mentioned effects of age group and the SNQ variables relationship strength and reciprocation on resource allocation, single- and multigroup models revealed effects of a number of additional variables. In the youngest network (midadolescent: Year 9), participants who scored higher on a measure of self-reported empathy (IRI EC) invested more in peers relative to those who scored lower on this measure (ceteris paribus), whereas this variable had no impact in older year groups (Table 5). The significant interaction between IRI EC and age group on investment in the multigroup model (Table 6) can be interpreted as indicating that the impact of this variable decreased with increasing year group. Although this result requires replication, younger participants were potentially utilizing an investment strategy based on their own feelings of empathy rather than on an appraisal of both self and other's views on each dyadic relationship. This interpretation is consistent with Selman's (1980) account of the development of perspective taking, in that midadolescents (aged 12–14 years) commonly engage in simple reciprocal perspective taking that takes into account self *or* other's perspective (potentially, one's own feelings of empathy), whereas older adolescents are more likely to simultaneously take into account self *and* other's perspectives on the relationship.

Another variable that exerted effects on giving in single and multigroup models was Duration 1 (whether participants had been mutually acquainted since primary vs. secondary school). Ceteris paribus, participants in the Year 10 network invested *less* in individuals to whom they had been acquainted since primary versus secondary school, and this effect was also evident in the multigroup model. That is, across the sample, participants gave less to peers to whom they had been acquainted for longer. However, since an analogous variable (duration of acquaintance in years) had no impact on investment in Study 1, this result requires replication before further conclusions can be drawn. Finally, Study 2 failed to find an effect of Machiavellian orientation (Kiddie Mach) on investment as observed in Study 1, either in single-group or multigroup models, which suggests that this finding may not be generalizable.

General Discussion

The current set of studies investigated development during adolescence in the impact of relationship reciprocation on resource allocation within established peer networks. For the first time, we show evidence for continuing development between mid- and late adolescence in the extent to which reciprocation of authentic social ties predicts prosocial cooperation behavior.

Across both studies, individual network regression (MRQAP) models were consistent with hypothesized age group differences in the impact of reciprocation. In midadolescence ($M_{\text{age}} = 14$), individuals invested more in peers to whom they reported stronger social ties (i.e., trust, liking, feelings of companionship). In later adolescence (ages = 15–17), individuals additionally invested more in peers who *reciprocated* strong ties. That is, late adolescents, but not midadolescents, were behaviorally sensitive to the degree to which others reciprocated their feelings of friendship, and adjusted their investment strategy accordingly.

In both studies, results from full-sample regression (MG-MRQAP) models showed that with increasing grade level, participants showed an increasing tendency to give more to classmates who reported a stronger relationship toward themselves than they reported toward their classmates. That is, between mid- and late adolescence (ages = 14–17), individuals' resource allocation decisions showed *increasing* sensitivity to the degree of relationship reciprocation. It is noteworthy that these convergent findings were obtained despite differences across studies in participant characteristics (private vs. state school; mixed- vs. single-gender networks), and differences in the mDG (e.g., use of real vs. fictitious money).

The current findings are consistent with prior developmental evidence from social-cognition and game theory paradigms suggesting that the tendency to take into account an interactant's perspective, motivations, and contributions in an experimental context continues to develop until at least midadolescence (Almås et al., 2010; Dumontheil et al., 2010; Eisenberg, Carlo, Murphy, & Van Court, 1995; Güroğlu et al., 2009; Van den Bos et al., 2010; Van Lange, Otten, De Bruin, & Joireman, 1997). In the current study, late adolescents, but not midadolescents, were able to use information regarding relationship reciprocation in resource allocation decisions. In contrast, midadolescent participants' resource allocation was determined by social tie strength (their own feelings of friendship) and (in Study 2) self-reported empathy. Potentially, older adolescents, with their more advanced perspective-taking abilities, were able to use information regarding their classmates' perceptions of the dyadic relationships to guide their investment decisions. In contrast, investment decisions of younger participants showed no evidence of sensitivity to the other's perspective on the relationship. These results are therefore also broadly consistent with the suggestion that social cognitive perspective taking shows protracted development during adolescence, particularly as assessed in the context of relationships with peers (Gurucharri & Selman, 1982; Selman, 1980). An important question is whether midadolescents were able to perceive relationship reciprocation subjectively but failed to use this information behaviorally (cf. Keysar, Lin, & Barr, 2003). Future studies are needed to unravel the direction of causality and to provide a broader framework within which to interpret the data. Studies should also investigate potential relations between the current findings and qualitative and quantitative observations of adolescent peer relationships (Bigelow & La Gaipa, 1980; Urberg et al., 1995; Youniss, 1982).

A number of additional points may be of note. Although the current studies investigated relationships within a number of school classes, in reality participants also have many relationships outside these classes. A common technique for investigating social networks based on peer nominations accounts for this complexity by asking each participant to name, for example, two close friends, and then asking these purported close friends to do the same, and so on (Marsden, 1990). However, although this technique has yielded important insights, it does not systematically map dyadic relationships bidirectionally and as such may be less interesting with regard to underlying social cognitive variables such as perspective taking. In contrast, because the method used in the current study maps dyadic ties exhaustively in a closed network (i.e., it maps each participant's relationship with each other participant), it provides information about bidirectional ties and thus reciprocation.

The current findings raise some interesting possibilities. First, maturation of behavioral sensitivity to reciprocity could contribute to important changes in peer relationships observed across adolescence. For example, a more nuanced understanding of bidirectional relationship qualities could contribute to developmental changes in expectations of friendships, and thus underpin observed decreases across adolescence in the number of self-reported close friendships (Urberg et al., 1995) and the increase in emphasis on intimacy and reciprocity (Youniss, 1982). Further studies utilizing the combination of experimental social-cognition tasks and mapping of real-world peer relationships may shed further light on this hypothesis. Second, the approach adopted here, which relates social network variables to an experimental measure of social behavior (Harrison et al., 2011), could potentially be utilized in neuroimaging research on dynamic social interactions and enduring social relationships, revealing general principles relating social cognitive and social network variables to their biological basis in the brain (Kanai, Bahrami, Roylance, & Rees, 2011; Lewis, Rezaie, Brown, Roberts, & Dunbar, 2011; Schilbach et al., 2013).

Finally, the current study illuminates a potentially important aspect of building positive social relationships, which are an important contributor to long-term mental health (House, Landis, & Umberson, 1988). Learning to invest in individuals who reciprocate prosocial behavior may promote more sustainable relationships, which in turn serve as a protective factor against mental illness. The ability to reciprocate cooperative behavior enables individuals to integrate as functioning members of social groups, while those who fail to do so may be punished and alienated (El Mouden, West, & Gardner, 2010; Fehr & Gächter, 2002). Understanding these sociocognitive processes during adolescence, which is a period of vulnerability to adverse mental health outcomes, could potentially be of value (Clifton, Pilkonis, & McCarty, 2007; Ormel et al., 2014; Paus, Keshavan, & Giedd, 2008).

Conclusions

The present set of studies used quantitative social network techniques to examine bidirectional relationship characteristics relevant for understanding subtle changes in cooperative behavior across adolescence. Results suggest development between mid- and late adolescence in the behavioral ability to recognize the degree to which others reciprocate feelings of friendship, in terms of impact on resource allocation in an mDG. These data underscore the importance of taking into account features of authentic relationships in research on adolescent prosocial behavior. Broadly, the mastery of sophisticated social strategies within adolescence may be critical for navigating an increasingly complex social world (Brown & Larson, 2009; Nelson et al., 2005). Combining social network methodology with experimental indices of prosocial behavior (Harrison et al., 2011) may reveal further insights into the development of strategic prosociality, its basis in the maturing adolescent brain, and its contribution to resilience across the life span.

References

Almås, I., Cappelen, A. W., Sørensen, E. Ø., & Tungodden, B. (2010). Fairness and the development of inequality acceptance. *Science*, *328*, 1176–1178. doi:10.1126/science.1187300

Bagwell, C. L., & Schmidt, M. E. (2013). *Friendships in childhood and adolescence*. New York, NY: Guilford Press.

Balliet, D., Li, N. P., Macfarlan, S. J., & Van Vugt, M. (2011). Sex differences in cooperation: A meta-analytic review of social dilemmas. *Psychological Bulletin*, *137*, 881–909. doi:10.1037/a0025354

Bigelow, B. J., & La Gaipa, J. J. (1980). The development of friendship values and choice. In H. C. Foot, A. J.

Chapman, & J. R. Smith (Eds.), *Friendship and social relations in children* (pp. 15–44). New York, NY: Wiley.

Borgatti, S., Everett, M., & Freeman, L. (2002). *Ucinet for windows: Software for social network analysis*. Harvard, MA: Analytic Technologies.

Brown, B. B., & Larson, J. (2009). Peer relationships in adolescence. In R. M. Lerner & L. Steinberg (Eds.), *Handbook of adolescent psychology* (Vol. 2., 3rd ed., pp. 74–103). New York, NY: Wiley. doi:10.1002/9780470479193.adlpsy002004

Burnett, S., & Blakemore, S.-J. (2009). The development of adolescent social cognition. *Annals of the New York Academy of Sciences*, *1167*, 51–56. doi:10.1111/j.1749-6632.2009.04509.x

Christie, R., & Geis, F. L. (1970). *Studies in Machiavellianism*. Waltham, MA: Academic Press.

Clifton, A., Pilkonis, P. A., & McCarty, C. (2007). Social networks in borderline personality disorder. *Journal of Personality Disorders*, *21*, 434–441. doi:10.1521/pedi.2007.21.4.434

Currie, C., Molcho, M., Boyce, W., Holstein, B., Torsheim, T., & Richter, M. (2008). Researching health inequalities in adolescents: The development of the Health Behaviour in School-Aged Children (HBSC) Family Affluence Scale. *Social Science & Medicine*, *66*, 1429–1436. doi:10.1016/j.socscimed.2007.11.024

Currie, C. E., Roberts, C., Morgan, A., Smith, R., Settertobulte, W., & Rasmussen, V. B. (2004). *Young people's health in context*. Copenhagen, Denmark: WHO Regional Office for Europe.

Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology*, *10*, 85.

Davis, M. H., & Franzoi, S. L. (1991). Stability and change in adolescent self-consciousness and empathy. *Journal of Research in Personality*, *25*, 70–87. doi:10.1016/0092-6566(91)90006-C

Dekker, D., Krackhardt, D., & Snijders, T. A. B. (2007). Sensitivity of MRQAP tests to collinearity and autocorrelation conditions. *Psychometrika*, *72*, 563–581. doi:10.1007/s11336-007-9016-1

Dumontheil, I., Apperly, I. A., & Blakemore, S.-J. (2010). Online usage of theory of mind continues to develop in late adolescence. *Developmental Science*, *13*, 331–338. doi:10.1111/j.1467-7687.2009.00888.x

Edele, A., Dziobek, I., & Keller, M. (2013). Explaining altruistic sharing in the dictator game: The role of affective empathy, cognitive empathy, and justice sensitivity. *Learning and Individual Differences*, *24*, 96–102. doi:10.1016/j.lindif.2012.12.020

Eisenberg, N., Carlo, G., Murphy, B., & Van Court, P. (1995). Prosocial development in late adolescence: A longitudinal study. *Child Development*, *66*, 1179–1197. doi:10.1111/j.1467-8624.1995.tb00930.x

Eisenberg, N., Fabes, R., & Spinrad, T. L. (2007). Prosocial development. *Handbook of child psychology* (pp. 646–718). Hoboken, NJ: Wiley.

El Mouden, C., West, S. A., & Gardner, A. (2010). The enforcement of cooperation by policing. *Evolution; International Journal of Organic Evolution*, *64*, 2139–2152. doi:10.1111/j.1558-5646.2010.00963.x

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, *415*, 137–140. doi:10.1038/415137a

Frank, S. A. (1998). *Foundations of social evolution*. Princeton, NJ: Princeton University Press.

Gardner, A., & Foster, K. R. (2008). The evolution and ecology of cooperation—History and concepts. In D. J. Korb & D. J. Heinze (Eds.), *Ecology of social evolution* (pp. 1–36). Berlin, Germany: Springer.

Goeree, J. K., McConnell, M. A., Mitchell, T., Tromp, T., & Yariv, L. (2010). The 1/*d* law of giving. *American Economic Journal: Microeconomics*, *2*, 183–203. doi:10.1257/mic.2.1.183

Güroğlu, B., van den Bos, W., & Crone, E. A. (2009). Fairness considerations: Increasing understanding of intentionality during adolescence. *Journal of Experimental Child Psychology*, *104*, 398–409. doi:10.1016/j.jecp.2009.07.002

Gurucharri, C., & Selman, R. L. (1982). The development of interpersonal understanding during childhood, preadolescence, and adolescence: A longitudinal follow-up study. *Child Development*, *53*, 924–927. doi:10.2307/1129129

Hammerstein, P., & Leimar, O. (2006). Cooperating for direct fitness benefits. *Journal of Evolutionary Biology*, *19*, 1400–1402. doi:10.1111/j.1420-9101.2006.01160.x

Harrison, F., Sciberras, J., & James, R. (2011). Strength of social tie predicts cooperative investment in a human social network. *PLoS ONE*, *6*, e18338. doi:10.1371/journal.pone.0018338

Hemelrijk, C. K. (1990). A matrix partial correlation test used in investigations of reciprocity and other social interaction patterns at group level. *Journal of Theoretical Biology*, *143*, 405–420. doi:10.1016/S0022-5193(05)80036-0

House, J. S., Landis, K. R., & Umberson, D. (1988). Social relationships and health. *Science*, *241*, 540–545. doi:10.1126/science.3399889

Kanai, R., Bahrami, B., Roylance, R., & Rees, G. (2011). Online social network size is reflected in human brain structure. *Proceedings of the Royal Society B: Biological Sciences*, *279*, 1327–1334. doi:10.1098/rspb.2011.1959

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition*, *89*, 25–41. doi:10.1016/S0010-0277(03)00064-7

Krackhardt, D. (1988). Predicting with networks: Nonparametric multiple regression analysis of dyadic data. *Social Networks*, *10*, 359–381. doi:10.1016/0378-8733(88)90004-4

Lewis, P. A., Rezaie, R., Brown, R., Roberts, N., & Dunbar, R. I. M. (2011). Ventromedial prefrontal volume predicts understanding of others and social network size. *NeuroImage*, *57*, 1624–1629. doi:10.1016/j.neuroimage.2011.05.030

Lorme, K. D., Bell, M. R., & Sisk, C. L. (2013). The teenage brain social reorientation and the adolescent brain —the role of gonadal hormones in the male Syrian hamster. *Current Directions in Psychological Science, 22,* 128–133. doi:10.1177/0963721413479607

Marsden, P. V. (1990). Network data and measurement. *Annual Review of Sociology, 16,* 435–463. doi:10.1146/annurev.so.16.080190.002251

Miyapuram, K. P., Tobler, P. N., Gregorios-Pippas, L., & Schultz, W. (2012). BOLD responses in reward regions to hypothetical and imaginary monetary rewards. *NeuroImage, 59,* 1692–1699. doi:10.1016/j.neuroimage.2011.09.029

Nachamie, S. S. (1970). Machiavellianism in children: The children's Mach Scale and the bluffing game. In R. Christie (Ed.), *Studies in Machiavellianism* (pp. 4550–4551). Waltham, MA: Academic Press.

Nelson, E. E., Leibenluft, E., McClure, E. B., & Pine, D. S. (2005). The social re-orientation of adolescence: A neuroscience perspective on the process and its relation to psychopathology. *Psychological Medicine, 35,* 163–174. doi:10.1017/S0033291704003915

Nettle, D., Colléony, A., & Cockerill, M. (2011). Variation in cooperative behaviour within a single city. *PLoS ONE, 6,* e26922. doi:10.1371/journal.pone.0026922

Nowak, M. A., & Sigmund, K. (1998). Evolution of indirect reciprocity by image scoring. *Nature, 393,* 573–577. doi:10.1038/31225

Ormel, J., Raven, D., van Oort, F., Hartman, C. A., Reijneveld, S. A., . . . Oldehinkel, A. J. (2014). Mental health in Dutch adolescents: A TRAILS report on prevalence, severity, age of onset, continuity and co-morbidity of DSM disorders. *Psychological Medicine, 20,* 1–16. doi:10.1017/S0033291714001469

Paus, T., Keshavan, M., & Giedd, J. N. (2008). Why do many psychiatric disorders emerge during adolescence? *Nature Reviews Neuroscience, 9,* 947–957. doi:10.1038/nrn2513

R Development Core Team. (2011). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing.

Rilling, J. K., & Sanfey, A. G. (2011). The neuroscience of social decision-making. *Annual Review of Psychology, 62,* 23–48. doi:10.1146/annurev.psych.121208.131647

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., & Bente, G., . . . & Vogeley, K. (2013). Toward a second-person neuroscience. *The Behavioral and Brain Sciences, 36,* 393–414. doi:10.1017/S0140525X12000660

Selman, R. L. (1980). *The growth of interpersonal understanding: Developmental and clinical analyses.* Waltham, MA: Academic Press.

Snijders, T., & Borgatti, S. (1999). Non-parametric standard errors and tests for network statistics. *Connections,* 22, 61–70.

Spear, L. P. (2000). The adolescent brain and age-related behavioral manifestations. *Neuroscience and Biobehavioral Reviews, 24,* 417–463. doi:10.1016/S0149-7634(00)00014-2

Spitzer, M., Fischbacher, U., Herrnberger, B., Grön, G., & Fehr, E. (2007). The neural signature of social norm compliance. *Neuron, 56,* 185–196. doi:10.1016/j.neuron.2007.09.011

Steinberg, L. (2008). A social neuroscience perspective on adolescent risk-taking. *Developmental Review, 28,* 78–106. doi:10.1016/j.dr.2007.08.002

Sutton, J., & Keogh, E. (2001). Components of Machiavellian beliefs in children: Relationships with personality. *Personality and Individual Differences, 30,* 137–148. doi:10.1016/S0191-8869(00)00017-9

Urberg, K. A., Değirmencioğlu, S. M., Tolson, J. M., & Halliday-Scher, K. (1995). The structure of adolescent peer networks. *Developmental Psychology, 31,* 540–547. doi:10.1037/0012-1649.31.4.540

Van den Bos, W., Westenberg, M., van Dijk, E., & Crone, E. A. (2010). Development of trust and reciprocity in adolescence. *Cognitive Development, 25,* 90–102. doi:10.1016/j.cogdev.2009.07.004

Van Lange, P. A., Otten, W., De Bruin, E. M., & Joireman, J. A. (1997). Development of prosocial, individualistic, and competitive orientations: Theory and preliminary evidence. *Journal of Personality and Social Psychology, 73,* 733–746. doi:10.1037/0022-3514.73.4.733

Wilson, D., Near, D., & Miller, R. (1996). Machiavellianism: A synthesis of the evolutionary and psychological literatures. *Psychological Bulletin, 119,* 285–299. doi:10.1037/0033-2909.119.2.285

Youniss, J. (1982). *Parents and peers in social development: A Sullivan-Piaget perspective.* Chicago, IL: University of Chicago Press.

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's website:

Figure S1. Study 1, Year 9 (Mid-Adolescent) Network. $N = 23$, M (SD) age = 14.46 (.26).

Figure S2. Study 1, Year 12 (late adolescent) Network. $N = 19$, M (SD) age = 17.22 (.29).

Figure S3. Study 2, Year 9 (Mid-Adolescent) Network. $N = 13$, M (SD) age = 14.10 (.36).

Figure S4. Study 2, Year 10 (Later Mid-Adolescent) Network. $N = 24$, M (SD) Age = 15.10 (.38).

Figure S5. Study 2, Year 11A (Late Adolescent) Network. $N = 25$, M (SD) Age = 15.96 (.24).

Figure S6. Study 2, Year 11B (Late Adolescent) Network. $N = 19$, M (SD) age = 16.00 (.28).

Appendix S1. Study 2 Network Data: t-Tests Comparing Pairs of Groups.

Child Development, September/October 2015, Volume 86, Number 5, Pages 1507–1521

On the Meanings of Self-Regulation: Digital Humanities in Service of Conceptual Clarity

Jeremy T. Burman
York University and The MEHRIT Centre

Christopher D. Green
York University

Stuart Shanker
York University and The MEHRIT Centre

Self-regulation is of interest both to psychologists and to teachers. But what the word means is unclear. To define it precisely, two studies examined the American Psychological Association's system of controlled vocabulary—specifically, the 447 associated terms it presents—and used techniques from the Digital Humanities to identify 88 closely related concepts and six broad conceptual clusters. The resulting analyses show how similar ideas are interrelated: self-control, self-management, self-observation, learning, social behavior, and the personality constructs related to self-monitoring. A full-color network map locates these concepts and clusters relative to each other. It also highlights some of the interests of different audiences, which can be described heuristically using two axes that have been labeled *abstract versus practical* and *self-oriented versus other-oriented.*

> . . . in psychology there are experimental methods *and conceptual confusion*. . . . The existence of the experimental method makes us think we have the means of solving the problems that trouble us; though problem and method pass one another by.
>
> Ludwig Wittgenstein (1953/2009),
> Philosophy of Psychology—A Fragment, §371

> Psychologists work with concepts that are often somewhat familiar to everybody—but the everyday terms are used in fuzzy and sloppy ways and carry lots of connotational baggage. Jargon is used because it is precise. New terms can be defined carefully, so that writers and informed readers share an exact understanding of what is meant.
>
> Roy Baumeister, quoted in "Words and Sorcery"
> (p. 198), by Oxenham and Sutton (2015)

There is a problem with self-regulation. The concept is extremely popular, but its definition is ambiguous: "management *of* the self, *by* the self" (e.g., Hammerberg, 2004, p. 366; Martin & McLellan, 2013, p. 135); "control *over* the self, *by* the self" (Audiffren & André, in press, p. 2; Baumeister & Exline, 2000, p. 30; Carlson, 2009, p. 88; Converse, Piccone, & Tocci, 2014, p. 65; Farley & Kim-Spoon, 2014, p. 434; Fischer, Greitemeyer, & Frey, 2007, p. 1309; Muraven & Baumeister, 2000, p. 247); and—self-referentially—"regulation *of* the self, *by* the self" (Bown & White, 2010, p. 434; Cukrowicz & Joiner, 2005, p. 160; Duckworth, 2011, p. 2639; Effeney, Carroll, & Bahr, 2013, p. 774; Flouri, Midouhas, & Joshi, 2014, p. 1044; Forgas, Baumeister, & Tice,

---

At the time of writing, both Burman and Shanker were jointly affiliated with The MEHRIT Centre, a nonprofit spin-off of the Milton and Ethel Harris Research Initiative at York University. Burman is now instead jointly affiliated with York University, Canada, and the University of Geneva (FPSE), Switzerland.

The research reported here was supported in part by grants from the Canadian Self-Regulation Initiative (to Shanker and Burman) and the Social Sciences and Humanities Research Council of Canada (to Green). Preliminary results were reported as part of a keynote address delivered by Burman at the CIRR National Reading Recovery & Early Literacy Conference: "Understanding Self-Regulation: What Is It? And Why Should You Care?" (February 14, 2014). The approach itself was inspired by the access granted to Burman during his service to the American Psychological Association (APA) on the board of advisors for PsycINFO (Electronic Resources Advisory Committee, 2007–2012), with new methods then developed subsequently in collaboration with Green's "PsyBorgs Lab" at York University. Thanks are thus due especially to Linda Beebe and Gary VandenBos, at APA, as well as Omar Alhassoon, Ellen Beckjord, Ian Galloway, Bradford Hesse, and Alvin Walker. The views reflected, however, are those of the authors.

Correspondence concerning this article should be addressed to Jeremy T. Burman, Department of Psychology, York University, Toronto, ON M3J 1P3, Canada. Electronic mail may be sent to jtburman@gmail.com.

© 2015 The Authors
Child Development © 2015 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2015/8605-0014
DOI: 10.1111/cdev.12395

2009, p. 4; Gawrilow et al., 2014, p. 806; Hofer, Busch, & Kärtner, 2011, p. 211; McDermott & Fox, 2010, p. 91; McKee, Ntoumanis, & Taylor, 2014, p. 300; Sheldon, Joiner, Pettit, & Williams, 2003, p. 305; punctuation, emphasis, and cited provenance varies). The result is that different interpretations are being reflected in the scholarly literature (cf. Dean, 1995, p. 563). Self-regulation, in other words, has *multiple discursive meanings* that are being acted upon as if they each were its sole scientific meaning (see also Boekaerts & Cascallar, 2006).

This is a challenge for psychological scientists because experimental findings can conflict needlessly as a result of clashing assumptions in setup, interpretation, and intent. (As is now well understood by philosophers of science, reported results are underdetermined by observation: they are "theory-laden" [following Hanson, 1958, pp. 2, 19; also Kuhn, 1962/2012]). The ambiguity is also a major hurdle for school teachers, especially those who want to apply ideas about self-regulation to improve their students' classroom experience, learning outcomes, and general well-being.

In short, the problem is this: in translating these findings across contexts for both specialist and public consumption, different implicit and unarticulated meta-theories will cause conflicts in interpreting the resulting *from knowing to doing* recommendations. But this then led us to wonder: *What else might be obscured by the lack of conceptual clarity?* (What might we be better able to understand—scientists and practitioners alike—if the meanings inherent to this important subject could be made more precise?)

To pursue this line of questioning specifically, we have taken advantage of the recent emergence of the Digital Humanities as a mode of scientific enquiry in psychology (e.g., Green, Feinerer, & Burman, 2013, 2014; Greenfield, 2013). Indeed, here, we have adopted some of its methodological strategies to parse the definitions informing the psychological study of self-regulation: we have taken advantage of a complete and agnostic third-party survey of the discipline's territories (the American Psychological Association's [APA] article-indexing program for PsycINFO), and used the resulting analyses to differentiate and focus on the different sets of meanings that shape psychologists' interpretations of observed phenomena. To make these meanings easier to see in relation to each other, we then also produced a relational network map (cf. Green, Feinerer, & Burman, 2015a, 2015b; Pettit, Serykh, & Green, 2015). This clearer perspective is intended to guide future research by signposting both the real and the rhetorical features of the discipline, while at the same time making the underlying concepts easier to understand, communicate, and apply.

Two studies are presented, examining the dictionary meanings of self-regulation as the concept is defined in psychology. The first study used the controlled vocabulary implemented in PsycINFO to identify the *dominant meanings* that influence American psychologists' perceptions. The second study then illustrates these findings—and locates the boundaries separating territories of potential conflict and collaboration (following Burman, 2009)—by showing how the dominant definitions cluster with lesser terms to form larger theory-laden meaning groups.

Our intent, broadly speaking, is to show the plurality of technical meanings that influence audiences' various understandings of self-regulation. But the resulting illustration is also akin to a web of belief (Quine & Ullian, 1970/1978). As a consequence, we are able not only to provide an overview of what it is *possible to say* about self-regulation in psychology, but also how different *partial* interpretations can conflict as strongly held visions of the same fundamental phenomena. In this, however, the extant conceptual imprecision represents not only a key translational issue: the problem is not just one of clearly communicating results so that they can be put into practice, but also one of clarifying the entirety of the set of potentially conflicting understandings that enable the pursuit of new discoveries as "thinkable things" (Lewis, 1943, pp. 238, 239, 240). Indeed, this is necessary if applications of the relevant concepts—otherwise indeterminate psychological and educational *gavagai* (Quine, 1960)—are to be accepted by skeptics as more than "just another educational fad" (Soupcoff, 2012, p. A14).

Study 1: Identifying the Dominant Senses of Self-Regulation

The semantics of psychological science—what the words used by psychologists mean to other psychologists—are governed by the *APA Dictionary* (VandenBos, 2006, 2009) and the *APA Thesaurus* (Tuleya, 2007). These sources provide the definitions that underpin the discipline. In this sense, they precede the understandings of those who would apply the insights arising from reported findings. Yet they are also the authority to which scientists and practitioners can turn in clarifying an ambiguous concept. To enable researchers to then find the relevant literature pertaining to the specific uses intended, stripped-down versions of the dictionaries have been integrated into PsycINFO (see "An Index Term," 2012; "Building Tools," 2012). The meanings they contain are accessible using a tool built directly into the APA's PsycNET interface: Term Finder.

PsycNET's Term Finder works just like a regular dictionary, except that its definitions are not provided in full sentence form (cf. "*of* the self, *by* the self"). Instead, meanings are implied by nested lists of related words: lists with entries defined by lists, which point to still more lists.

Here, we have captured the relations defined by these lists to examine what their relatedness implies about the meanings of the interconnected concepts. In other words, we have constructed a network: we have mapped self-regulation's definitional web. To present the results in a way that could then be used as an intuition pump—a tool for thought to inform new science, new practice, and new policy (Dennett, 2013)—we have also used a free and open-source visualization package to illustrate the data: Gephi (Bastian, Heymann, & Jacomy, 2009).

Gephi's purpose is to convert relational data into network maps. It can also be used to conduct analyses that describe these maps in quantitative terms. In our first study, however, we were not interested so much in *describing* the contents of the dictionary as we were concerned with *identifying the most influential meanings* that the APA provides for self-regulation. As a result, what follows diverges somewhat from standard network analyses: rather than starting with an illustration of everything and its interconnections, we instead implemented a blind link analysis to deduce the source concept's primary meanings. These then provided a set of foci that we used in interpreting the subsequent visualization. And they allowed us to go beyond the standard discussions of network analyses (e.g., regarding topographical centrality), so we could stay focussed on our problem: presenting a clearer definition of self-regulation. (To jump ahead to the two new definitions we have abstracted from our analyses, see the large blocks of italicized text at the end of the article.)

Method

Each entry in the controlled vocabulary is linked-to in Term Finder. We captured these links by defining them using two columns in a spreadsheet program: the entry itself was noted in a column labeled *source* and each of the definition-terms was noted in a new row of a column labeled *target*. Every unique entry in either column thus became interpretable by the software as a node in our network, and the relations defined along the rows across the columns became the network's connecting edges. (For a clear but technical discussion of how such an approach can be used to identify communities from the interconnections thus defined, see Girvan & Newman, 2002.)

We started with the Self Regulation entry, which is the term in the APA's controlled vocabulary that most closely matches our chosen concept of self-regulation. (To make it clear when we are referring to a controlled term, these will be referred to in caps with no hyphens.) This process of defining edges using dictionary entries as nodes was then repeated for all listed terms, down through three levels of definitions, to capture all of the terms that are relatable to self-regulation (and without regard for our own particular subject-related research interests).

For those who wish to replicate the work, the procedure was as follows: (0) identify a term in the controlled vocabulary that describes the concept you seek to define (i.e., for self-regulation, we chose Self Regulation) and record it in the source column of the spreadsheet. With the term identified, then: (1) capture the source's relations to all of its associated terms—which Term Finder provides under headings titled "used for," "broader," "narrower," and "related"—by recording each of them in the target column in a new row while at the same time carrying over the source so that every row is populated by two entries; (2) for each of the associated terms, repeat these first steps to capture what we will call their *second-order definitions* by setting the associated term as the new source and each second-order definition as a target in its own row; (3) for each of those second-order definitions, capture what we will call their *third-order anchors* by again noting sources and targets in new rows; and (4) remove duplicates, so that repeated mentions of the same relations are not misinterpreted by the software as being quantitatively meaningful.

The resulting list will contain entries that refer to each other, and it is this cross-referencing that is most important in terms of defining the network topography: two sources that refer to the same target are related by their shared reference to it, and all of their associated terms are also thus related to each other indirectly by that connection. We then exported these data from the spreadsheet and imported them into the Data Laboratory workspace in Gephi. To precisely characterize the influence of each definition in the overall network, and thereby identify the primary meanings of self-regulation, we used the tools provided in the Statistics panel to calculate each node's PageRank (Brin & Page, 1998).

PageRank provides the mathematical basis for Google's search engine. It is their principal method of determining the relevance of any one document in a large collection of mutually interlinking texts (e.g., the pages of the World Wide Web). However, we used it to provide a rank ordering of our identified terms and quantify the distance between them.

The system works by providing a measure of connectedness, reflecting the observation that highly cited sources contribute more to the perceived meaning of a collection than less well-known texts. (More correctly, however, PageRank is a probabilistic model of findability if a searcher were to randomly land on any one node in the network and then seek to find another node solely by traveling the available links: a greater number of inbound links implies greater relevance because those nodes are more findable from more places in the network.) Following this approach, individual nodes can then be ranked according to how they are linked-to by all of the other nodes: a finite amount of findability (100%) is apportioned throughout the entire network, with more findable nodes ranking more highly.

The standard empirical article would at this point present inferential statistics as results. That is not necessary here: our sampling of the relevant contents of the dictionary is complete. However, the PageRank itself is indeed a calculated probabilistic metric. For this reason, the *z* score is a convenient tool for reporting conceptually significant differences in findability. This also provides the means to derive the standard measures with which psychologists are familiar, and thus the *z* score offers a useful rhetorical device for explaining our results (viz. using *z*-critical values, for given α, thereby providing a nonparametric equivalent to the more common p values). In addition, the results are then comparable between different analyses of different terms of interest because the scaling will always be relative to the topology of each term's defining network: when the total amount of a given network's findability is always only 100%, comparing the primary meanings of concepts with different networks is greatly simplified.

Results

According to the APA's controlled vocabulary, the meaning of the Self Regulation index term is interpretable through the definitions of seven related terms: Agency, Emotional Regulation (used for affect regulation), Self Control (used for willpower), Self Management, Self Monitoring (used for self observation), the parenthetically named Self Monitoring (Personality), and Self Regulated Learning. Each of these is then interpretable through their own sets of related terms, expanding the total number of implicated definitions to 49. And the meanings of this second-order set can be anchored, topographically, such that the total number of terms considered is 447—all of which are distinct and locatable. (This anchoring is crucial for the method of visualizing meanings used in Study 2: terms linked-to by other relevant definitions will be drawn in closer to each other by the layout algorithm, while terms not linked-to will be pushed outward to the periphery, preserving topographical centrality and proximity as simple heuristics in interpreting meanings.)

Together, the three levels define all aspects of everything connected in the *APA Dictionary* to self-regulation as a scientific concept. If something has not been captured by this method, it is simply because the term has not (yet) been defined in that way by the APA.

The PageRank analysis produces a rank ordering of connectedness with a very large number of low-relevance nodes and a very small number of high-relevance nodes. These results can be seen directly for top terms in Table 1, while the histogram in Figure 1 simplifies all of the data and shows their skew. What this all means, though, is fairly simple: if a randomly walking searcher were to start at a series of different locations in the network, and follow different paths while exploring the conceptual landscape, then they would likely pass in predictable proportions through the same small number of highly connected points.

That said, however, this analysis also shows something surprising about the meaning of self-regulation: the most highly connected term in the network is not Self Regulation (with a *z* score of 6.4)—as we would have expected, given that this is where we started—but rather Self Control (11.9). Indeed, as we can see from the *z* scores listed in Table 1, this dominant definition is almost twice as findable as the expected term ($\alpha < .001$).

This surprising result is a function of the interrelatedness of terms in the dictionary: more terms associated with self-regulation point to Self Control as one of their meanings than to anything else. That said, however, Self Control (11.9) is not alone in outranking Self Regulation (6.4) in findability. It is

Table 1

Self-Regulation Terms With Unfiltered z Scores > 0.5 (i.e., Top 44 of 447 Possible)

	Study 1		Study 2	
Controlled term	PageRank (unfiltered)	z score (unfiltered)	Cluster	Other boundaries
Self Control	1.12	11.86	Self Control	
Self Monitoring	0.78	7.43	Self Monitoring/Self Management	
Self Regulation	0.70	6.38	N/A	Self Monitoring/Self Management Self Control
Self Management	0.68	6.03	Self Monitoring/Self Management	
Emotional Regulation	0.67	6.00	Self Control	
Self Monitoring (Personality)	0.54	4.21	Self Monitoring (Personality)	
Emotional Control	0.48	3.34	Self Control	
Self Evaluation	0.47	3.27	Self Monitoring/Self Management	
Self Perception	0.40	2.41	Self Monitoring (Personality)	
Agency	0.40	2.35	Agency/Self Determination/ Locus of Control/Helplessness	
Self Determination	0.39	2.28	Agency/Self Determination/ Locus of Control/Helplessness	
Behavior Modification	0.39	2.27	Self Monitoring/Self Management	
Internal External Locus of Control	0.37	1.93	Agency/Self Determination/ Locus of Control/Helplessness	Self Control
Helplessness	0.35	1.63	Agency/Self Determination/ Locus of Control/Helplessness	
Self Regulated Learning	0.34	1.54	Self Control	Learning/Learning Strategies
Personality Traits	0.34	1.48	Self Monitoring (Personality)	
Affect Regulation [not shown]	0.33	1.46	Self Control	Indexed as Emotional Regulation ($z = 6.00$)
Reflectiveness	0.32	1.29	Self Monitoring (Personality)	
Self Report	0.32	1.25	Self Monitoring/Self Management	
Anger Control	0.32	1.22	Self Monitoring/Self Management	Self Control
Self Help Techniques	0.31	1.20	Self Monitoring/Self Management	
Impulse Control Disorders	0.31	1.15	Self Control	Social Behavior
Personality	0.30	1.08	Self Monitoring (Personality)	
Cognitive Therapy	0.30	1.08	Self Monitoring/Self Management	
Personality Processes	0.30	1.04	Self Monitoring (Personality)	Social Behavior Self Monitoring/Self Management
Self Criticism	0.30	1.03	Self Monitoring (Personality)	Self Monitoring/Self Management
Introspection	0.30	0.97	Self Monitoring (Personality)	
Impulsiveness	0.29	0.89	Self Control	Social Behavior
Behavioral Disinhibition	0.29	0.87	Self Control	Social Behavior
Volition	0.29	0.87	Agency/Self Determination/ Locus of Control/Helplessness	
Interpersonal Control	0.29	0.85	Agency/Self Determination/ Locus of Control/Helplessness	
Social Comparison	0.29	0.82	Self Monitoring/Self Management	
Social Perception	0.29	0.82	Self Monitoring/Self Management	
Self Efficacy	0.28	0.75	Agency/Self Determination/ Locus of Control/Helplessness	Self Control Self Monitoring/Self Management
Cognitive Techniques	0.28	0.69	Self Monitoring/Self Management	
Learning	0.27	0.66	Learning/Learning Strategies	
Empowerment	0.27	0.65	Agency/Self Determination/ Locus of Control/Helplessness	
Perfectionism	0.27	0.62	Self Monitoring (Personality)	

Table 1
Continued

Controlled term	Study 1		Study 2	
	PageRank (unfiltered)	z score (unfiltered)	Cluster	Other boundaries
Temptation [not shown]	0.27	0.61	Agency/Self Determination/ Locus of Control/Helplessness	Near to Individuality/Independence (Personality)
Control (Self) [not shown]	0.27	0.61	Self Control	Indexed as Self Control (z = 11.86)
Willpower [not shown]	0.27	0.61	Self Control	Indexed as Self Control (z = 11.86)
Learning Strategies	0.27	0.59	Learning/Learning Strategies	
Self Instructional Training	0.27	0.58	Self Monitoring/Self Management	
Psychotherapeutic Techniques	0.26	0.52	Self Monitoring/Self Management	
Reinforcement	0.26	0.50	Learning/Learning Strategies	Self Monitoring/Self Management

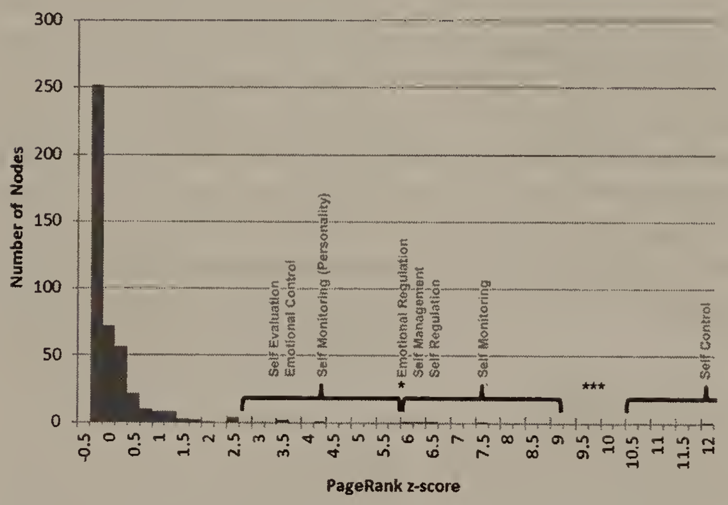


Figure 1. Scaled comparison of the different "findabilities" in the unfiltered network defining self-regulation (Study 1). It is clear that some terms are significantly more findable (more relevant) than others. Top terms are labeled directly in the histogram; asterisks indicate the degree of difference, in terms of scaled distance, from the central term.

also joined by Self Monitoring (7.4). Yet while this is itself significantly different from Self Control ($\alpha < .001$), it is not so from Self Regulation ($\alpha > .05$). In fact, in examining the *z* scores we see that the four top terms defining self-regulation after Self Control are all approximately equivalent in their influence: Self Monitoring (7.4), Self Regulation (6.4), Self Management (6.0), and Emotional Regulation (6.0) are all significantly different from the next highest ranked term—Self Monitoring (Personality), with a *z* score of 4.2—but they are not significantly different from each other ($\alpha > .05$). This blurring of secondary meanings has important implications, which we will discuss and develop throughout the remainder of this article.

Discussion

The sum of all of the PageRanks for all 447 nodes is 1. In other words, tautologically, if a searcher starts at all nodes, then the probability of their finding any node is 100%. (The whole, in this case, is identical with the sum of its parts.) The challenge is then to reduce the number of parts to a meaningful representative sample: which of the 447 nodes are *most relevant* to our problem? (What concepts would need to be included in a clearer definition?)

If Google's method—calculating findability—is acceptable as a means of determining relevance (in more than a colloquial sense), then focusing on the highest ranking nodes can provide an assessment of the most relevant contributors to the meaning of the overall network. From this perspective, then, the *primary* meaning of self-regulation is therefore that of self-control: if this were a Google search, then Self Control is the node in this network that would receive the most referred traffic (i.e., nearly double that of the next most relevant node). Yet the next four terms—the network's secondary meanings—would *collectively* receive more traffic still. Therefore, the meaning of this network as a whole cannot be so simply defined: not only is the present meaning of self-regulation ambiguous, but *so is what it is possible to meaningfully say about self-regulation*.

That said, however, we cannot end there. It is not clear from the PageRank ordering alone whether what we seek is most appropriately attached to the individual definitions, or if they ought to be treated in clusters that collect together groups of definitions to form higher order concepts.

That will be examined in Study 2, with a view to ultimately providing a new—clearer—definition for self-regulation.

Conclusion From Study 1

Noting the relations between the terms defined in the APA's controlled vocabulary enabled us to delve more deeply into the various meanings of a source term: Self Regulation. Rather than having to rely on a list of seven potentially ambiguous dictionary-provided definitions, however, we were able to use how these terms are themselves defined by the APA to access hundreds of related interpretations. These were then ranked using a standard algorithm (Google's PageRank), providing an ordered list of the most influential associated meanings.

In other words, by taking advantage of techniques more often associated with search and information retrieval, we have been able to treat the APA's controlled vocabulary as a directed network. From this perspective, definitions point outward from a source term and inward toward their associated target definitions (i.e., the direction from source to target is what operationalizes the association and expresses one term's relatedness to the other). Analyzing the connectedness of these links then provided us with a quantitative estimation of the influence of the most influential meanings used by the discipline's dominant professional association in parsing and interpreting the science of self-regulation.

The primary result was surprising, but it is also consistent with the public's recent embracing of popular books on this topic: the public understanding of self-regulation does indeed seem to be that of willpower (Baumeister & Tierney, 2012; McGonigal, 2012), focus (Goleman, 2013), and grit (Duckworth, in press; Tough, 2012). These are all stories of self-control (see also Mischel, 2014). Yet this does not seem to us to be as useful—to teachers—as some of the alternative meanings that are obscured by the influence of the self-control narrative, such as Self Monitoring, Self Management, and Emotional Regulation. (Briefly: If it is possible for a child to achieve the goal of greater self-control by learning and practicing skills related to self-observation and self-management, then this suggests a different approach to controlling disruptive classroom behavior than that typically taken today.) The next logical step, then, is to adopt a broader perspective and see if the view is different; to expand from the consideration of influential single terms to an examination of the broader discourses to which they contribute.

Study 2: The Discourses of Self-Regulation

The standard approach in network analysis is to identify and then examine communities of similarity (e.g., Green et al., 2015a, 2015b). This is often given a sociological cast (e.g., Pettit et al., 2015). Yet our data have been derived from the APA's controlled vocabulary rather than from how those meanings have been used by the authors who collectively make up the discipline: our study is one of *semantics*, not *pragmatics*. (Due to limitations of space, and the added complexity of the methods required to extract these additional data from PsycINFO, we will save that more explicitly social examination for a follow-up study.) Here, therefore, we want simply to define the discourses *in which it is possible* for communities of experts to engage. That will then provide a more solid foundation upon which to build in future work.

Our approach in Study 2 was to pursue the goal identified in Study 1: to reduce the number of parts, while at the same time preserving what matters most of the whole. We did this by identifying clusters of meanings, using Gephi's modularity tool (see Blondel, Guillaume, Lambiotte, & Lefebvre, 2008). Then, because the probability of finding a cluster during a search can be understood as being a function of the probability of finding its members, we needed only to sum together the findabilities of the cluster members using the PageRanks calculated in Study 1. That then provided an assessment of the influence of each of the clusters on the overall meaning of self-regulation as a scientific concept.

Study 1 examined relations between terms to determine findability (a formal approach for defining relevance), but it did not visualize the results. Study 2 therefore provides the visualization expected of network analysis, but then we leverage the results from Study 1 to extend its implications still further.

Method

Picking up directly from Study 1, in Gephi, we switched from the Data Laboratory workspace to the Overview workspace. We then used the Force Atlas 2 layout algorithm to organize the nodes according to the interrelations defined by their edges (see Jacomy, Venturini, Heymann, & Bastian, 2014). This provided an unfiltered relational cloud.

But it also included many definitions that are irrelevant to our intent: we wanted to focus solely on those terms that matter most to self-regulation as a scientific concept, while ignoring those that do not. (This filtering required a further step, taken after others that required the raw data.)

With the data mapped, we used the modularity tool to identify clusters of meaning. So that the resulting analysis would reflect the interconnectedness in the dictionary, we turned on the *use edge weights* option. We also used the default resolution of 1. The resulting modularity of the clusters was calculated to be 0.550. Repeated analyses changed the boundaries slightly, but the affected nodes were small and the resulting differences minor. (Still, these minor differences are accommodated below in the discussion of boundary objects.)

The modularity analysis' groupings informed our group-findability calculations: the PageRank results from Study 1 were summed according to their calculated group membership. These calculations are inferential (they are based on both the PageRanks and the modularity analysis), but the network itself is not. That is a key point: the network is an illustration of the relations defined in the controlled vocabulary. The semantic map is therefore itself a description, not an inference.

Using the in-degree filter, we hid all nodes with fewer than two inward definition-seeking links. ("In-degree" is a more local measure than PageRank [discussed by Fortunato, Boguñá, Flammini, & Menczer, 2008].) This then filtered out all of the least individually relevant nodes of the network, and reduced the number shown in the visualization from 447 to the 88 most conceptually significant terms. However, we preserved the original unfiltered layout provided by the algorithm: the effect of these less meaningful nodes, in the visualization, is to pull unrelated terms further apart while allowing connected terms to remain close to each other. That distance, in the visualization, is intuitively valuable: nodes that are closer together are typically also more closely related.

Results

Gephi's modularity analysis identified six clusters at the default resolution. To simplify the resulting map, the members of each cluster have been color-coded (see Figure 2, with legend and associated data in Table 2). To simplify naming, they have also been labeled according to their dominant nodes (using the *z* scores calculated in Study 1). Finally, to distinguish between node names and cluster names, we will use **bold text** in referring to the clusters.

Listing the clusters clockwise, starting from the top-right, provides the following territories in our map of the semantics of self-regulation:

1. **Learning/Learning Strategies**
2. **Self-Monitoring/Self Management** (including Self Evaluation and Behavior Modification)
3. **Agency/Self Determination/Internal External Locus of Control/Helplessness**
4. **Self-Control** (including Emotional Regulation and Emotional Control)
5. **Social Behavior**
6. **Self-Monitoring (Personality)** (including Self Perception, Personality Traits, Reflectiveness, and Personality)

To make the map easier to read, we have also added some boundaries. These have been placed at the approximate midpoints between the groups, where the clusters blur into each other. Orphans that have been separated from their cluster—or that sit on boundaries—are contained in circles to separate and mark them out specifically.

We think that this map provides a reasonable parsing of the conceptual meanings associated with self-regulation as a scientific concept. The large clusters include enough members to afford a useful simplification, the vertices separating them are intuitively sensible, and the overall organization makes sense: **Self Monitoring (Personality)** on the left transitions through both **Self Monitoring/Self Management** and **Self Control** (via **Social Behavior**) in the center before arriving at **Agency/Self Determination/Locus of Control/Helplessness** on the lower-right. And the Self-Regulated Learning node is located, as expected, with the **Learning/Learning Strategies** cluster.

From these results, and taking advantage of our earlier calculation of each node's PageRank (in Study 1), it was a simple matter to calculate the findability of each cluster. Table 2 summarizes these data. The key finding is given by the column labeled Percentage of Map (filtered).

Self Monitoring/Self Management is thus shown to account for 27.3% of the meaning of self-regulation and **Self Monitoring (Personality)** for slightly less with 21.8%. **Learning/Learning Strategies** is also better represented than expected with 17.7%, while **Self Control**—surprisingly—accounts for only 16.5%. **Agency/Self Determination/Locus of Control/Helplessness** accounts for 15.9%, and **Social Behavior** for barely 1% (although it is worth

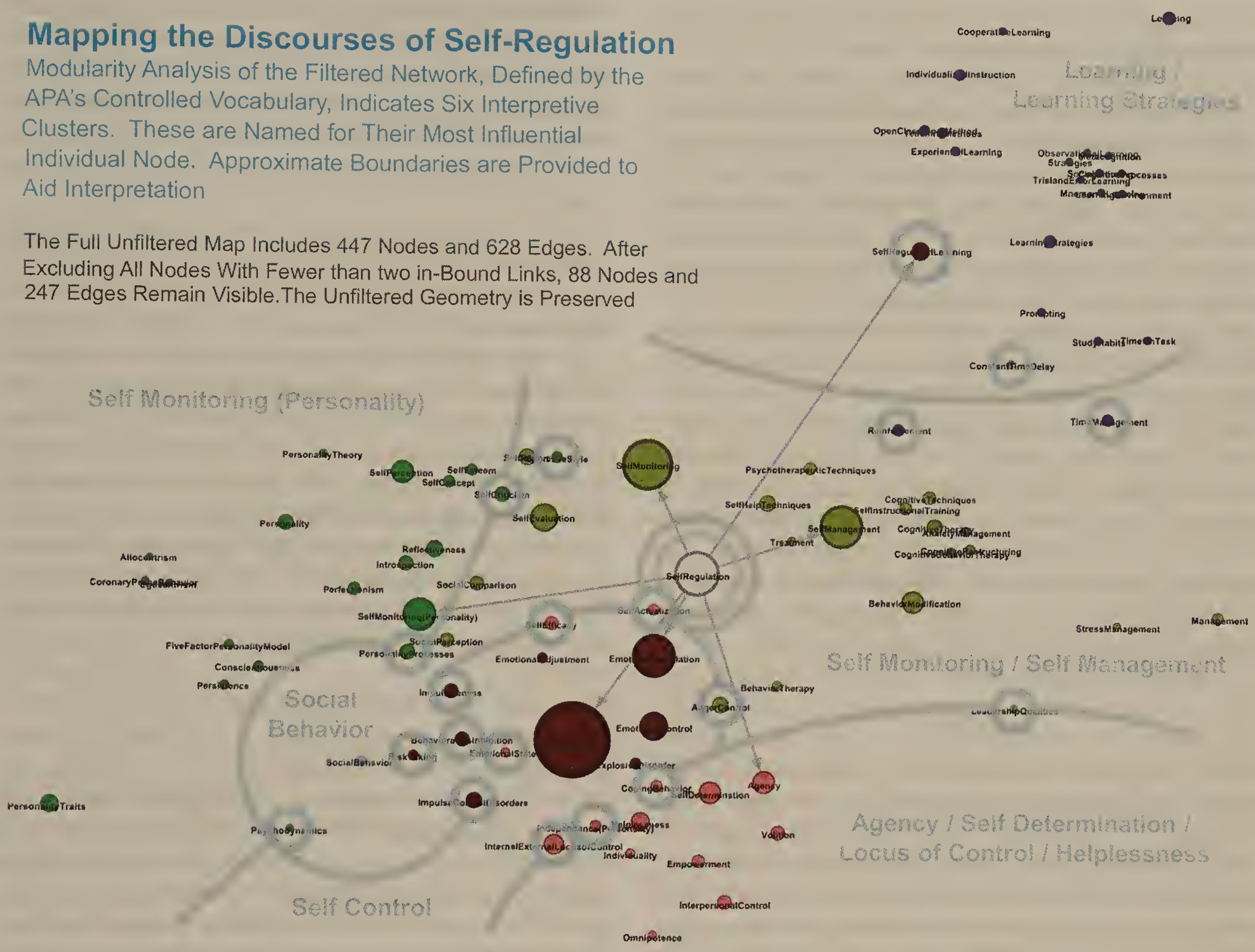

Mapping the Discourses of Self-Regulation
Modularity Analysis of the Filtered Network, Defined by the APA's Controlled Vocabulary, Indicates Six Interpretive Clusters. These are Named for Their Most Influential Individual Node. Approximate Boundaries are Provided to Aid Interpretation
The Full Unfiltered Map Includes 447 Nodes and 628 Edges. After Excluding All Nodes With Fewer than two in-Bound Links, 88 Nodes and 247 Edges Remain Visible.The Unfiltered Geometry is Preserved
Learning / Learning Strategies
Self Monitoring (Personality)
Self Monitoring / Self Management
Social Behavior
Agency / Self Determination / Locus of Control / Helplessness
Self Control
CooperativeLearning
IndividualizedInstruction
ExperientialLearning
SelfRegulatedLearning
LearningStrategies
Prompting
ConstantTimeDelay
TimeManagement
PsychotherapeuticTechniques
SelfHelpTechniques
SelfManagement
CognitiveTechniques
Treatment
BehaviorModification
SelfRegulation
SelfMonitoring
StressManagement
Management
BehaviorTherapy
PersonalityTheory
SelfPerception
SelfEsteem
SelfConcept
SelfEvaluation
Personality
Allocentrism
Reflectiveness
Introspection
Perfectionism
SocialComparison
SelfMonitoring(Personality)
SelfEfficacy
SocialPerception
FiveFactorPersonalityModel
Conscientiousness
Persistence
EmotionalAdjustment
BehavioralInhibition
SocialBehavior
ImpulseControlDisorders
EmotionalControl
Agency
SelfDetermination
Volition
InternalExternalLocusofControl
Individuality
Empowerment
InterpersonalControl
Omnipotence
PersonalityTraits
Psychodynamics

Figure 2. Mapping the discourses of self-regulation.

Table 2
Cluster Analysis as Map Legend (Color-Coded)

Modularity class	Color code on map	Name	Sum of PageRanks (unfiltered)	Percentage of map (filtered) (%)	Rank	Number of nodes	Density per term on map
−1	White	Self Regulation	N/A	N/A		1	N/A
0	Maroon	Self Control	4.21	16.48	4	10	1.65
1	Pink	Agency/Self Determination/ Locus of Control/Helplessness	4.06	15.89	5	14	1.13
2	Army Green	Self Monitoring/Self Management	6.97	27.29	1	22	1.24
3	Green	Self Monitoring (Personality)	5.56	21.77	2	20	1.09
4	Light Blue	Social Behavior	0.23	0.92	6	1	0.92
5	Dark Blue	Learning/Learning Strategies	4.51	17.65	3	20	0.88

noting that *the social* is implicated in terms grouped with other clusters, such as the nearby nodes Social Perception and Social Comparison).

To help readers focus on the most significant nodes, in considering the individual terms separately from the clusters to which they belong, these have each been sized according to their calculated PageRank: the larger the node, the more findable it is. The cluster-findability scores can then be understood to be a sum of the areas of the nodes in that color code (although similarities between colors are not themselves meaningful).

Finally, it is worth noting that we have expanded the boundary in the visualization for the **Social Behavior** cluster to overlap more strongly with its neighbors than we initially thought to do. This is to reflect the recognized intersection between Social and Personality, as subdisciplinary domains of psychology, and also our pragmatic understanding of the nearby concepts. (For example: risk taking is more likely among teens when they are in a peer group, but less likely when with their parents [Farley & Kim-Spoon, 2014; Rai et al., 2003].) Indeed, it seems more theoretically appropriate to label this group "co-regulation." Yet this would require its own argument (e.g., Hadwin & Oshige, 2011; Volet, Summers, & Thurman, 2009). Thus, to keep things simple—even though the resulting labeling is not as scientifically useful as we might like—we have kept to the labels provided by the *z* scores.

Discussion

The use of Gephi's in-degree filter is crucial for simplifying the visualization of the data generated by Study 1. Only those definitions with more than one referring entry in the controlled vocabulary have been included in the map (Figure 2). This reduced the number of visible nodes by more than 80%, and yet only reduced the number of visible edges by 60%. In other words, what we might call the *information density* of the remaining mapped nodes has been doubled by the filter's use: we have retained only 20% of the terms offered by the dictionary (nodes) but accounted for 40% of its provided definitions (edges).

Delving into this more deeply, it is clear that the filter eliminated a large number of definitions from the controlled vocabulary that are not relatable to self-regulation in a scientifically meaningful way. For example: the **Learning/Learning Strategies** cluster of self-regulation is defined relevantly in the map as being concerned with Meta Cognition, Learning Environment, and Experiential Learning. Yet, without the filter, it was also defined irrelevantly as involving Rat Learning, Cat Learning, and Machine Learning. (These are specialist subject areas that may indeed involve discussions of self-regulation, but then—because articles are described with more than one index term from the controlled vocabulary—those specific pragmatic uses would be captured by other relevant terms.) Similarly, relevant traits associated with **Self Monitoring (Personality)** include Persistence and Conscientiousness, while others like Masculinity and Machiavellianism are irrelevant. (Thus, also with **Social Behavior**, Social Media is filtered out.) These irrelevant meanings are not shown on the map, and their omission represents an improvement in signal quality.

As in Study 1, a primary meaning has been identified (see Table 2). Yet this is not as expected: the dominant concept, after the cluster analysis, is **Self Monitoring/Self Management**. Indeed, between this and **Self Monitoring (Personality)**, nearly half of the map can be accounted for (49.1%). The primary meaning expected from the results of Study 1—**Self Control**—ranks fourth between **Learning/Learning Strategies** and **Agency/Self Determination/Locus of Control/Helplessness**. This apparent contradiction between studies is peculiar: *How could one term from the controlled vocabulary account for so much of the dictionary meaning of our target term when the broader concept it supposedly represents accounts for so little?*

Conclusion From Study 2

In this second study we have been able to go beyond the consideration of individual terms to examine what large groupings of relevant terms contribute to self-regulation's conceptual topography. This then affords the basis for a much more precise definition of the term than is presently accepted, and requires only a little more space: *Self-regulation involves the monitoring and management of the self, by the self, so that behavior can be appropriately controlled—especially in learning situations. Differences in personality traits are implicated, as are differences in individual agency (or, more popularly, in willpower). But self-monitoring dominates: the control of the self is preceded by observations of the self. There is also a social component, but its effects are typically captured by other means.*

This reflects the most relevant majority of what it is possible to say about self-regulation, according to the APA's definition of the related terms. Yet it also reflects the ambiguity in the meaning of underlying concepts—especially "self" (cf. Martin & McLellan, 2013, p. 135). Still, though, because it is becoming increasingly evident that it is inappropriate to treat such phenomena as being wholly innate (as if there were a gene for willpower), this definition could be pushed further so that the language is more consistent with more recent and emerging developmental, dynamic, epigenetic, and systemic perspectives. This would represent an extension of the concepts reviewed, however, because those terms are not reflected in the map; it requires a meta-theoretical rejection of a nativist theory-laden view of self-control in considering self-regulation (see the next large block of italicized text below).

General Discussion of Both Studies, Presenting a New Definition

From 447 related concepts extracted from the APA's controlled vocabulary, we have identified 88 relevant variations: the dimensions of self-regulation, according to how the discipline's dominant professional association defines the associated terms. We then reduced this set to the six broad conceptual meanings that organize all of the implicated terms. In other words, we have confirmed that the meaning of self-regulation is ambiguous, but in formalizing its discourses we now also aim to make new progress simpler and easier to achieve (following Burman, 2009).

Beyond the six meanings, there seem also to be some broad axes at work in organizing the map: left to right from **Self Monitoring (Personality)** to **Self Monitoring/Self Management** decomposes almost along abstract-applied lines, while top to bottom (or, perhaps seen more easily in the mind's eye, far to close) from **Learning/Learning Strategies** to **Self Control** decomposes as other-oriented versus self-oriented (with an embedding in social interaction at **Social Behavior**). From this perspective, the central concept that we sought to define in the first place—Self Regulation—is shown to be at the middle of it all: everything to everyone, and thus not very helpful when considered on its own.

Implications for Knowledge Translation

These broad axes give us the opportunity to frame contributions in a new way. We can even attribute ideal professional personae to the six meanings of self-regulation, which may then help scholars in their attempts to orient to the different audiences that can be understood to adopt these conceptual selves as norms (cf. Morawski, 2007). Thus, for example, we can suggest that those interested in **Learning/Learning Strategies** (e.g., teachers) are likely to seem extremely other-oriented and extremely practical simply by virtue of their concern with the topics grouped with this cluster. By contrast, those interested in **Self Monitoring (Personality)** are likely to seem only moderately self-oriented yet extremely abstract (i.e., interested in personality assessment). Those interested in **Agency**, extremely self-oriented and moderately practical (in the sense that their approach will typically be more abstract than what teachers are apt to ask for). **Self Control** then reflects a midpoint between the two self-oriented perspectives (which perhaps explains some of its popular success), with those interested in aspects related to **Social Behavior** seeming somewhat more abstract and somewhat less practical. This leaves **Self Monitoring/Self Management**, the primary meaning for self-regulation suggested by Study 2: ideally interested in both self-orientation and other-orientation, while tending toward the practical but still accepting of some abstractions.

Of course, the view out from inside any one cluster will be relative to its perspective. Distance on the map can thus be understood as a proxy for epistemic distance, with personal attributions in response to criticism then made somewhat more predictable: a distant other will seem either self-absorbed or a martyr, condescending or naïve, and irrelevant or missing the point. Such evaluations speak more about positionality than the content of the claims, however, and are thus to be treated gently: they are not criticisms, but opportunities for clarification and growth.

Still, the six meanings are not straightforward in their implications. They blur together. After adding topographical boundaries to simplify the resulting map, we also then identified several nodes that require special attention. We will call them "boundary objects" (following Star & Griesemer, 1989). These have implications for knowledge translation too, since they mark territories of conflicting interests (delivering on Burman, 2009; Kitto, Sargeant, Reeves, & Silver, 2012).

Boundary objects

Boundary objects can be adapted to a variety of perspectives, according to the interests of those referring to them. We have identified several candidates, and marked them on the map with a circle: Self Regulated Learning, Constant Time Delay, Reinforcement, Time Management, Leadership Qualities, Anger Control, Coping Behavior, Independence (Personality), Internal External Locus of Control, Impulse Control Disorders, Emotional States, Behavioral Inhibition, Risk Taking, Impulsiveness, Personality Processes, Self Criticism, Cognitive Style, Self Efficacy, and Self Actualization. These are all concepts that share multiple meanings, in the sense that they seem to flip back and forth between interpretations, and thus require the most care in discussion and—especially—in translation from knowing to doing. Self Regulation itself, obviously, is one too: not only does it share boundaries with **Self Monitoring/Self Management** and **Self Control**, but it also alludes to all of the other meanings that we have mapped.

To try to accommodate the impact of these boundary objects on the meanings with which they are associated, we have recalculated the cluster-PageRanks for the map (Table 3). With one exception, very little changes in the cluster findabilities. But **Social Behavior** jumps by 300%. It is thus more than *implicated* in other topics (e.g., Social Perception and Social Comparison). Rather, it might be more accurate to say that its relevance and importance has been *obscured* by how we conceive of this material (cf. Burman, 2015; also in Hobbs & Burman, 2009).

Related to this, we are also now able to point out something curious about the language of psychology itself: when accounting for the number of nodes involved in defining each of these clusters, we see that what we have called their "information density" is lowest in **Learning/Learning Strategies** (0.9% per term) and highest in **Self Control** (1.6%–1.7% per term; compare Tables 2 and 3). This suggests that there is greater descriptive precision in the language of learning, at least when considering its connection to self-regulation, and much less precision in the language of self-control. This then perhaps also explains the latter's apparent dominance when fewer terms are taken into account: according to Zipf's Law, the rank order of a term is directly proportional to the frequency of its use (see Zipf, 1935). The skew toward Self Control that we identified in Study 1 could thus be illusory: a function solely of its relative imprecision, and thus also of the availability heuristic (Tversky & Kahneman, 1973).

Recognizing this—and also that it is primarily the self-control discourse driving the association with innate ability (viz. willpower, focus, grit)—we are now able to push our earlier empirically derived definition further: *Learning self-regulation involves learning how to monitor and manage one's internal states and predispositions, so that behavior can be appropriately controlled not only in the classroom but also in any other situation where one is expected to become better educated or informed. This developing faculty is measurable using the standard tools of the study of individual differences, but especially those involved in studying personality. Not only is the Will implicated, but so too is the developmental pathway that can increase its power and effectiveness in controlling behavior. In other words: the learning of self-regulatory skills, especially those related to self-observation and self-management, has implications for how the Will manifests and is then measured (with further downstream effects). Or, rather: From skill comes capacity. There is also a social component to self-regulation, but in psychological studies its influence is most often captured indirectly by the measurement of other factors.* In short: the notion of self-regulation is much more complex than the "management," "control," or "regulation *of* the self, *by* the self." Indeed, it encapsulates much of what psychological thinkers have been struggling with for the last several hundred years: all three domains of the neoclassical psyche—cognition, affect, and conation—are implicated (see Hilgard, 1980; also Berrios & Gili, 1995).

The challenge, for future research, will be to see how these in-theory dictionary-style definitions are actually used by research psychologists. (Has conceptual imprecision, and availability bias, skewed research in self-regulation away from potentially useful approaches?) But we leave that examination of use-meanings (pragmatics) for later, because there are complex methodological issues that need

Table 3

The Six Selves of Self-Regulation, With Cluster Percentages Adjusted to Account for Boundary Objects (Area-Based)

Name	Sum of PageRanks (unfiltered)	Percentage of map (filtered) (%)	Rank	Delta versus Table 2 (%)	Number of nodes	Density per term on map
Self Control	4.72	17.96	3	8.98	11.16	1.61
Agency/Self Determination/ Locus of Control/Helplessness	3.26	12.40	5	−21.94	11.84	1.05
Self Monitoring/Self Management	7.78	29.61	1	8.50	24	1.23
Self Monitoring (Personality)	4.98	18.97	2	−12.84	18.17	1.04
Social Behavior	1.00	3.81	6	316.28	3.83	0.99
Learning/Learning Strategies	4.53	17.24	4	−2.33	19	0.91

to be dealt with first. Those preliminaries are in preparation, however, as the effects of the Digital Humanities revolution continue to spread into psychology (e.g., Priva & Austerweil, 2015).

Final Concluding Thoughts

There is presently no precise, general, and widely accepted definition for self-regulation. There are many; dozens of competing perspectives with slightly different emphases. Yet any single scientific study can address only a handful. When results are communicated, authors therefore need to be clear about which of our six broad conceptual foundations their work seeks to build upon: **Learning/Learning Strategies, Self-Monitoring/Self Management, Agency/Self Determination/Locus of Control/Helplessness, Self Control, Social Behavior**, or **Self-Monitoring (Personality)**. These areas overlap, of course, but the boundaries between them represent areas of both future collaboration and potential misunderstanding.

It is heartening, however, that the most topographically central term in our map—apart from Self Regulation itself—is Self Actualization (a member of the **Agency** cluster, but located at the boundary between **Self Monitoring/Self Management** and **Self Control**). This is also reassuring to us personally because it is consistent with the motivation driving our translational work: we hope to foster the recognition and encouragement of *whatever it is* that makes it easier for children to achieve the goals they can learn to set for themselves (including, but not limited to, those of the classroom).

The colloquial expression for this broad goal is *autonomy*, and the associated controlled term is Independence (Personality). The formal semantic meaning of this boundary object, in relation to self-regulation, is then captured by the collection of concepts grouped under **Self Control** and **Agency/Self Determination/Locus of Control/Helplessness**. This is a direct reflection of the term's definition in the controlled vocabulary, which presents it as one of many concepts related broadly to Personality Traits (**Self Monitoring [Personality]**) and specifically to Empowerment (**Agency**), Internal External Locus of Control (**Agency**), Resistance (**Agency** [not shown]), and Self Determination (**Agency**). But that is not exactly what we mean either.

Our hope beyond identifying its centrality in the meaning of self-regulation is also to foster its development in situations and environments where **Learning/Learning Strategies** is the focus, while also encouraging the development of affiliated skills related to **Self Monitoring/Self Management** and affiliative co-regulatory **Social Behavior**. In other words, we do not mean to encourage autonomous self-actualizing in the narrow sense; we intend it in a way that reflects facets of the entire map. And, indeed, a classroom influenced by such an approach would be a very different place: less focused on meting out rewards and punishments to control behavior, and more focused on providing co-constructive interactions and nudges that help lead the children to produce similar or better outcomes more easily and with less fuss. (For related readings, see Shanker, 2012; Shanker & Barker, in press; Sokol, Grouzet, & Müller, 2013.)

References

An index term by any other name . . . (2012). *PsycINFO News*, *31*(1), 3–5.

Audiffren, M., & André, N. (in press). The strength model of self-control revisited: Linking acute and chronic effects of exercise on executive functions. *Journal of Sport and Health Science*. Advance online publication. doi:10.1016/j.jshs.2014.09.002

Bastian, M., Heymann, S., & Jacomy, M. (2009). *Gephi: An open source software for exploring and manipulating networks*. International AAAI Conference on Weblogs and Social Media. Retrieved from http://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154/1009

Baumeister, R. F., & Exline, J. J. (2000). Self-control, morality, and human strength. *Journal of Social and Clinical Psychology*, *19*, 29–42. doi:10.1521/jscp.2000.19.1.29

Baumeister, R. F., & Tierney, J. (2012). *Willpower*. London, UK: Allen Lane/Penguin.

Berrios, G. E., & Gili, M. (1995). Will and its disorders: A conceptual history. *History of Psychiatry*, *6*, 87–104. doi:10.1177/0957154X9500602105

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment, P10008*. doi:10.1088/1742-5468/2008/10/P10008

Boekaerts, M., & Cascallar, E. (2006). How far have we moved toward the integration of theory and practice in self-regulation? *Educational Psychology Review*, *18*, 199–210. doi:10.1007/s10648-006-9013-4

Bown, J., & White, C. (2010). Affect in a self-regulatory framework for language learning. *System*, *38*, 432–443. doi:10.1016/j.system.2010.03.016

Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. [Proceedings of the Seventh International World Wide Web Conference]. *Computer Networks and ISDN Systems*, *30*, 107–117. doi:10.1016/S0169-7552(98)00110-X

Building tools for efficient searching: Updating the thesaurus of psychological index terms®. (2012). *PsycINFO News*, *31*(1), 1–4.

Burman, J. T. (2009). Convergent plurality or basic incommensurability? (Toward the formalizing of Goertzen's solution to the "crisis" in psychology). *History and Philosophy of Psychology Bulletin, 20*, 23–28.

Burman, J. T. (2015). Neglect of the foreign invisible: Historiography and the navigation of conflicting sensibilities. *History of Psychology, 18*, 146–169. doi:10.1037/a0039194

Carlson, S. M. (2009). Social origins of executive function development. In C. Lewis & J. I. M. Carpendale (Eds.), *Social interaction and the development of executive function* (pp. 87–97). San Francisco, CA: Jossey-Bass. doi:10.1002/cd.237

Converse, P. D., Piccone, K. A., & Tocci, M. C. (2014). Childhood self-control, adolescent behavior, and career success. *Personality and Individual Differences, 59*, 65–70. doi:10.1016/j.paid.2013.11.007

Cukrowicz, K. C., & Joiner, T. E., Jr. (2005). Treating the "mischances of character," simply and effectively. *Journal of Contemporary Psychotherapy, 35*, 157–168. doi:10.1007/s10879-005-2697-9

Dean, M. (1995). Governing the unemployed self in an active society. *Economy and Society, 24*, 559–583. doi:10.1080/03085149500000025

Dennett, D. C. (2013). *Intuition pumps and other tools for thinking*. New York, NY: Norton.

Duckworth, A. L. (2011). The significance of self-control. *Proceedings of the National Academy of Sciences of the United States of America, 108*, 2639–2640. doi:10.1073/pnas.1019725108

Duckworth, A. L. (in press). *Grit: Passion, perseverance and the science of success*. New York, NY: Scribner.

Effeney, G., Carroll, A., & Bahr, N. (2013). Self-regulated learning and executive function: Exploring the relationships in a sample of adolescent males. *Educational Psychology, 33*, 773–796. doi:10.1080/01443410.2013.785054

Farley, J. P., & Kim-Spoon, J. (2014). The development of adolescent self-regulation: Reviewing the role of parent, peer, friend, and romantic relationships. *Journal of Adolescence, 37*, 433–440. doi:10.1016/j.adolescence.2014.03.009

Fischer, P., Greitemeyer, T., & Frey, D. (2007). Ego depletion and positive illusions: Does the construction of positivity require regulatory resources? *Personality and Social Psychology Bulletin, 33*, 1306–1321. doi:10.1177/0146167207303025

Flouri, E., Midouhas, E., & Joshi, H. (2014). Family poverty and trajectories of children's emotional and behavioural problems: The moderating roles of self-regulation and verbal cognitive ability. *Journal of Abnormal Child Psychology, 42*, 1043–1056. doi:10.1007/s10802-013-9848-3

Forgas, J. P., Baumeister, R. F., & Tice, D. M. (2009). The psychology of self-regulation: An introductory review. In J. P. Forgas, R. F. Baumeister, & D. M. Tice (Eds.), *Psychology of self-regulation: Cognitive, affective, and motivational processes* (pp. 1–17). New York, NY: Psychology Press.

Fortunato, S., Boguñá, M., Flammini, A., & Menczer, F. (2008). Approximating PageRank from in-degree. In W. Aiello, A. Broder, J. Janssen, & E. Milios (Eds.), *Algorithms and models for the web-graph* (pp. 59–71). Berlin, Germany: Springer. [Lecture notes in computer science, no. 4936]. doi:10.1007/978-3-540-78808-9_6

Gawrilow, C., Fäsche, A., Guderjahn, L., Gunzenhauser, C., Merkt, J., & von Suchodoletz, A. (2014). The impact of self-regulation on preschool mathematical achievement. *Child Indicators Research, 7*, 805–820. doi:10.1007/s12187-013-9201-y

Girvan, M., & Newman, M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America, 99*, 7821–7826. doi:10.1073/pnas.122653799

Goleman, D. (2013). *Focus*. New York, NY: HarperCollins.

Green, C. D., Feinerer, I., & Burman, J. T. (2013). The new schools of psychology 1: Digital analysis of *Psychological Review*, 1894–1903. *Journal of the History of the Behavioral Sciences, 49*, 167–189. doi:10.1002/jhbs.21592

Green, C. D., Feinerer, I., & Burman, J. T. (2014). Beyond the schools of psychology 2: Digital analysis of *Psychological Review*, 1904–1923. *Journal of the History of the Behavioral Sciences, 50*, 249–279. doi:10.1002/jhbs.21665

Green, C. D., Feinerer, I., & Burman, J. T. (2015a). Searching for the structure of early American psychology: Networking *Psychological Review*, 1894–1908. *History of Psychology, 18*, 15–31. doi:10.1037/a0038406

Green, C. D., Feinerer, I., & Burman, J. T. (2015b). Searching for the structure of early American psychology: Networking *Psychological Review*, 1909–1923. *History of Psychology, 18*, 196–204. doi:10.1037/a0039013

Greenfield, P. M. (2013). The changing psychology of culture from 1800 through 2000. *Psychological Science, 24*, 1722–1731. doi:10.1177/0956797613479387

Hadwin, A., & Oshige, M. (2011). Self-Regulation, coregulation, and socially shared regulation: Exploring perspectives of social in self-regulated learning theory. *Teachers College Record, 113*, 240–264.

Hammerberg, D. D. (2004). Technologies of the self in classrooms designed as "learning environments": (Im) possible ways of being in early literacy instruction. In B. M. Baker & K. E. Heyning (Eds.), *Dangerous coagulations: The uses of Foucault in the study of education* (pp. 359–382). New York, NY: Peter Lang.

Hanson, N. R. (1958). *Patterns of discovery: An inquiry into the conceptual foundations of science*. Cambridge, UK: Cambridge University Press.

Hilgard, E. R. (1980). The trilogy of mind: Cognition, affection, and conation. *Journal of the History of the Behavioral Sciences, 16*, 107–117. doi:10.1002/1520-6696(198004)16:2 < 107::AID-JHBS2300160202 > 3.0.CO;2-Y

Hobbs, S., & Burman, J. T. (2009). Is the "cognitive revolution" a myth? *The Psychologist, 22*, 812–815.

Hofer, J., Busch, H., & Kärtner, J. (2011). Self-regulation and well-being: The influence of identity and motives. *European Journal of Personality, 25*, 211–224. doi:10.1002/per.789

Jacomy, M., Venturini, T., Heymann, S., & Bastian, M. (2014). ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLoS ONE*, *9*, e98679. doi:10.1371/journal.pone.0098679

Kitto, S., Sargeant, J., Reeves, S., & Silver, I. (2012). Towards a sociology of knowledge translation: The importance of being dis-interested in knowledge translation. *Advances in Health Science Education*, *17*, 289–299. doi:10.1007/s10459-011-9303-6

Kuhn, T. S. (2012). *The structure of scientific revolutions* (50th anniversary ed.). Chicago, IL: University of Chicago Press. (Original work published 1962)

Lewis, C. I. (1943). The modes of meaning. *Philosophy and Phenomenological Research*, *4*, 236–250. doi:10.2307/2103080

Martin, J., & McLellan, A.-M. (2013). *The education of selves: How psychology transformed students*. New York, NY: Oxford University Press. doi:10.1093/acprof:oso/9780199913671.001.0001

McDermott, J. M., & Fox, N. A. (2010). Exploring response monitoring: Developmental differences and contributions to self-regulation. In R. H. Hoyle (Ed.), *Handbook of personality and self-regulation* (pp. 91–113). Chichester, UK: Wiley-Blackwell. doi:10.1002/9781444318111.ch5

McGonigal, K. (2012). *The willpower instinct: How self-control works, why it matters, and what you can do to get more of it*. New York, NY: Avery/Penguin.

McKee, H. C., Ntoumanis, N., & Taylor, I. M. (2014). An ecological momentary assessment of lapse occurrences in dieters. *Annals of Behavioral Medicine*, *48*, 300–310. doi:10.1007/s12160-014-9594-y

Mischel, W. (2014). *The marshmallow test: Mastering self-control*. New York, NY: Little, Brown/Hachette.

Morawski, J. G. (2007). Scientific selves: Discerning the subject and the experimenter in experimental psychology in the United States, 1900–1935. In M. G. Ash & T. Sturm (Eds.), *Psychology's territories: Historical and contemporary perspectives from different disciplines* (pp. 129–148). Mahwah, NJ: Erlbaum.

Muraven, M., & Baumeister, R. F. (2000). Self-regulation and depletion of limited resources: Does self-control resemble a muscle? *Psychological Bulletin*, *126*, 247–259. doi:10.1037/0033-2909.126.2.247

Oxenham, S., & Sutton, J. (2015). Words and sorcery. *The Psychologist*, *28*, 198–201.

Pettit, M., Serykh, D., & Green, C. D. (2015). Multispecies networks: Visualizing the psychological research of the Committee for Research in Problems of Sex. *Isis*, *106*, 121–149. doi:10.1086/681039

Priva, U. C., & Austerweil, J. L. (2015). Analyzing the history of *Cognition* using topic models. *Cognition*, *135*, 4–9. doi:10.1016/j.cognition.2014.11.006

Quine, W. V. O. (1960). *Word and object*. Cambridge, MA: MIT Press.

Quine, W. V. O., & Ullian, J. S. (1978). *The web of belief* (2nd ed.). New York, NY: Random House. (Original work published 1970)

Rai, A. A., Stanton, B., Wu, Y., Li, X., Galbraith, J., Cottrell, L., & Burns, J. (2003). Relative influences of perceived parental monitoring and perceived peer involvement on adolescent risk behaviors: An analysis of six cross-sectional data sets. *Journal of Adolescent Health*, *33*, 108–118. doi:10.1016/S1054-139X(03)00179-4

Shanker, S. (2012). *Calm, alert, and learning: Classroom strategies for self-regulation*. Toronto, Canada: Pearson Education Canada.

Shanker, S. & Barker, T. (in press). *Self-reg*. New York, NY: Penguin.

Sheldon, K. M., Joiner, T. E., Jr., Pettit, J. W., & Williams, G. (2003). Reconciling humanistic ideals and scientific clinical practice. *Clinical Psychology: Science and Practice*, *10*, 302–315. doi:10.1093/clipsy.bpg026

Sokol, B. W., Grouzet, F. M. F., & Müller, U. (Eds.). (2013). *Self-regulation and autonomy: Social and developmental dimensions of human conduct*. New York, NY: Cambridge University Press. doi:10.1017/CBO9781139152198

Soupcoff, M. (2012, September 25). The fad mentality. *National Post*, p. A14.

Star, S. L., & Griesemer, J. R. (1989). Institutional ecology, "translations" and boundary objects: Amateurs and professionals in Berkeley's museum of vertebrate zoology, 1907–39. *Social Studies of Science*, *19*, 387–420. doi:10.1177/030631289019003001

Tough, P. (2012). *How children succeed*. Boston, MA: Houghton Mifflin Harcourt.

Tuleya, L. G. (Ed.). (2007). *Thesaurus of psychological index terms*. Washington, DC: American Psychological Association.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*, 207–232. doi:10.1016/0010-0285(73)90033-9

VandenBos, G. R. (Ed.). (2006). *APA dictionary of psychology*. Washington, DC: American Psychological Association.

VandenBos, G. R. (Ed.). (2009). *APA concise dictionary of psychology*. Washington, DC: American Psychological Association.

Volet, S., Summers, M., & Thurman, J. (2009). High-level co-regulation in collaborative learning: How does it emerge and how is it sustained? *Learning and Instruction*, *19*, 128–143. doi:10.1016/j.learninstruc.2008.03.001

Wittgenstein, L. (2009). Philosophy of psychology—A fragment. In G. E. M. Anscombe, P. M. S. Hacker, & J. Schulte (Trans.) & P. M. S. Hacker & J. Schulte (Eds.), *Philosophical investigations* (4th Rev ed., pp. i-v, 189-192, vii-ix, 197-198, x, 200-202, xi, 204-240, xii-xiv; formerly known as "Part II"). Malden, MA: Wiley-Blackwell. (Original work published 1953)

Zipf, G. K. (1935). *The psycho-biology of language*. Boston, MA: Houghton Mifflin.

Child Development, September/October 2015, Volume 86, Number 5, Pages 1522–1537

Expectations About Ethnic Peer Group Inclusivity: The Role of Shared Interests, Group Norms, and Stereotypes

Aline Hitti
Tulane University

Melanie Killen
University of Maryland

This study investigated three factors that contribute to social exclusion: group norms, individual characteristics, and stereotypes. Non-Arab American 12- and 16-year-olds ($N = 199$) judged their expectations about the inclusivity of Arab American and non-Arab American peer groups toward new peers characterized by: (a) different ethnic identity but similar interests (e.g., hobbies) and (b) same ethnic identity but different interests. Participants expected that when groups had exclusive norms, Arab American peers would base inclusion decisions on ethnic identity, but that their own non-Arab group would base decisions on shared interests. Participants who reported stereotypes expected their in-group to be ethnically less inclusive. With age, ethnic-based exclusion increased. The findings are discussed in light of current research on developmental intergroup relationships.

Given the increase in mobility of people around the world, children today are encountering peers from a wide range of different ethnic backgrounds. For children who are members of an ethnic minority group, and who may already feel like outsiders, extensive experiences of social exclusion impede healthy adjustment to school environments (Stone & Han, 2005). Research has demonstrated that social exclusion based on ethnic membership is different from peer rejection based on individual characteristics, such as temperament, shyness, and social deficits (Killen, Mulvey, & Hitti, 2013). Unlike peer rejection based on individual characteristics, ethnic-based social exclusion has to do with the role of prejudice and bias as well as the unfairness experienced by children who are excluded because of their gender, race, and ethnicity. Research over the past decade has shown that stereotypes, prejudicial attitudes, and negative biases account for some of the peer-based social exclusion in childhood. Other research has also documented the instances in which children and adolescents reject ethnic-based exclusion, thus capturing their moral concerns about the injustice associated with prejudice and discrimination (for a review, see Killen & Rutland, 2011).

The social reasoning developmental (SRD) perspective (Rutland, Killen, & Abrams, 2010) has provided a framework for conducting research on social exclusion in intergroup contexts. This model integrates social identity development theories (Nesdale, 2004; Tajfel, 1979) and social domain theory (Turiel, 1983) to guide research on social exclusion. SRD integrates these theories to make predictions about developmental changes in children's inclusive and exclusive orientations as a function of moral concerns, group identity, group norms, intergroup contact, perceived out-group threat, and stereotypic assumptions (Rutland et al., 2010). Therefore, when evaluating exclusion of others, children try to balance between competing concerns for fairness and being inclusive as well as threats to group identity or cohesion. What factors are given priority change developmentally as identification with one's own ethnic group emerges and understanding of group dynamics (i.e., how groups work) increases with age. Consequently,

---

This research was funded by the College of Education, University of Maryland, Support Program for Advancing Research and Collaboration (SPARC) awarded to the first author. We thank Kelly Lynn Mulvey, Shelby Cooley, Laura Elenbaas, Michael Rizzo, and Jeeyoung Noh for their invaluable input on this project. In addition we want to thank the undergraduate research assistants who helped with data collection, coding, and analysis: Sofia Castro, Samantha Lichbach, Jamie Ott, Phillip Jordan, Kaye Schacter, Amy Komen, Benjamin White, Hillary Johnson, Naina Vohra, Nilo Fallay-Sohy, Leon Li, and Chandni Patel. Importantly, we would like to thank the schools and students who participated in this study.

Correspondence concerning this article should be addressed to Aline Hitti, Department of Psychology, Tulane University, 2007 Percival Stern Hall, New Orleans, LA 70118. Electronic mail may be sent to ahitti@tulane.edu.

© 2015 The Authors
Child Development © 2015 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2015/8605-0015
DOI: 10.1111/cdev.12393

children's intergroup judgments can be influenced by both the norms of the groups (moral or conventional) and attitudes about out-groups. Additionally, the emergence of stereotypes about out-groups plays a role in whether children rely on fairness or intergroup attitudes to make group decisions about inclusion. The goal of the current study was to understand how group norms and stereotypes are related to the extent to which children give priority to individual characteristics when making interethnic group inclusion decisions.

One of the pervasive findings to emerge has been that children who reject or exclude peers because of their group membership (gender, race) provide reasons for their actions that are based on "group functioning," which is often defined as the perceived problem that arises from including someone who is "different" into a group (Killen, Lee-Kim, McGlothlin, & Stangor, 2002). What is not well understood is what "being different" means when used as an explanation to justify peer exclusion. Research on peer relationships shows the choice of friendships or inclusion of peers into one's social group is based on shared interests, values, and beliefs (Rubin, Bukowski, & Parker, 2006; Stark & Flache, 2012). However, Stark and Flache (2012) showed that peer friendships based on shared opinion (i.e., opinion homophily) can be disguised as ethnic homophily if ethnicity is highly correlated with opinion (e.g., "African Americans don't like pop music" or "Latinos do not like school"). Stereotypic assumptions made about an ethnic out-group's interests, values, and beliefs can therefore interfere with the extent to which an ethnic majority child might view peers from an ethnic minority as similar.

While some research has shown that young children give priority to shared interests over race when provided with the opportunity for intergroup contact (see McGlothlin & Killen, 2005, 2010), little is known about whether children older than 7–10 years with little out-group contact would focus on shared interests over group membership, especially when evaluating group decisions about including new peers into their friendship groups. Targeting older children is important given findings from an SRD perspective (Rutland et al., 2010) showing adolescents more frequently reference "group functioning" to justify exclusion than do younger children. Adolescents also have greater experience with groups than younger children (Brown, Eicher, & Petrie, 1986) and thus are expected to have a more sophisticated understanding of group dynamics. Extending the focus on shared interests and group membership to the context of inclusion and exclusion for children, as well as adolescents, who have very little contact with an ethnic out-group, was a central goal of this study. When shared interests are given priority in ethnic intergroup encounters, this implies that children are making ethnically inclusive choices.

Another goal of this study was to test the role of exclusive and inclusive peer group norms in whether children and adolescents focused on shared interests or ethnic group membership. In a study by Nesdale, Maass, Durkin, and Griffiths (2005), children as young as 9 years of age, who were experimentally assigned to groups with exclusive norms (e.g., your group does not like to work with children in the other group, especially if they are different) preferred out-group members less than those assigned to groups with inclusive norms. Findings from the same line of research also show that if schools support inclusive norms then the effects of exclusive peer group norms are mitigated (Nesdale & Lawson, 2011). Moreover, children's perceptions that an ethnic out-group is inclusive predicts their willingness to partake in cross-ethnic friendships (Tropp, O'Brien, & Migacheva, 2014). Attention to group norms and the type of group norms continues through adolescence, impacting intergroup attitudes (Mähönen, Jasinskaja-Lahti, & Liebkind, 2011) as well as reasoning about exclusion and group dynamics (Abrams, Rutland, Pelletier, & Ferrell, 2009; Hitti, Mulvey, Rutland, Abrams, & Killen, 2014). Thus, examining whether children's and adolescents' judgments about the use of shared interests or group membership varies as a function of inclusive or exclusive peer group norms was investigated in the current study.

Additionally, assessing participants' expectations about inclusion decisions made by both their own group and an ethnic out-group provided further insight into the obstacles that are created when a majority group perceives a minority group to be exclusive. Past research has assessed expectations about groups based on long histories of national rivalries (e.g., Abrams et al., 2009) or gender (Mulvey, Hitti, Rutland, Abrams, & Killen, 2014). The current study extended the research in a new direction by focusing on a target out-group of U.S. children of Arab descent. This group was chosen because most U.S. children and adolescents have had little contact with children of Arab background, and yet this is a group associated with negative stereotypes, especially given the events in the past decade in the United States, post 9/11 (Ibish, 2008). This context is one that has been occurring in many places around the globe as communities migrate to new countries to seek asylum, such as Muslims in the Netherlands (Verkuyten, 2008), or to take refuge from civil wars and strife, such as with ex-Yugoslavians in Switzerland (Malti, Killen, & Gasser, 2012).

Arab Peers as an Ethnic Out-Group in the U.S. Context

According to Naber (2000), who extensively discussed the paradoxes of the Arab American identity, the first wave of Arab immigrants to the Unites States in the late 1800s included Arabs of Christian faith who assimilated into the "White" Christian American majority. Recent Arab immigrants in the past decade (2000s), who mostly ascribe to the Muslim faith, have had more difficulty integrating in American culture. Ethnic majority U.S. children typically have little contact with Muslim children of Arab descent. Societal norms that condemn exclusion of Arabs are weak, just as many norms regarding immigrants in Europe, which is also currently dealing with tensions due to an influx of new Muslim and Arab immigrants. In addition, in the United States, Arab as an ethnic out-group category is an understudied group in developmental research but is an important one to consider given that Arabs have been stigmatized because of associations with terrorism and negative portrayals in the media (Shaheen, 2003).

Recent research in the United States has shown that children and individuals of Arab descent are at times treated differently by their peers due to their religion or language (Flanagan, Syvertsen, Gill, Gallay, & Cumsille, 2009). Research has documented examples of misconceptions and stereotypes about Arab Americans (e.g., "They are all Muslim," "They are racist") as well as a high rate of prejudice and discrimination aimed at this group (DeRosier, 2004; Flanagan et al., 2009; Ibish, 2008). What has not been examined is whether children and adolescents use stereotypes to make inclusion and exclusion decisions about Arab out-group peers. The current study investigated the effects of stereotypes, in the form of generalizations made about an Arab out-group, given that stereotypes have been shown to influence social judgments in the absence of other individuating information (Horn, 2003; Killen et al., 2002).

The Current Study

This study had four central goals. The first goal was to determine age-related differences regarding whether non-Arab American children and adolescents gave priority to group membership (ethnic identity) or individual characteristics (shared interests) when deciding who peer groups should include into the group. The second goal was to test whether group norms (inclusive or exclusive) factored into inclusion or exclusion decisions. The third goal was to examine the role of group identity in children's understanding of group dynamics by determining whether participants differed when judging expectations about groups' exclusivity for the in-group (non-Arab American) or the out-group (Arab descent). The fourth goal was to examine whether stereotypic expectations were related to expectations about ethnic out-group inclusivity and exclusivity.

In the current study, participants were asked to anticipate how groups would make inclusion decisions regarding same-gender peer groups that differed by ethnic identity (Arab American and non-Arab American). To examine whether children expected peer groups to give priority to shared interest over ethnic identity, two target peers were identified as wanting to join the group and they had the following characteristics: (a) different ethnic identity but does share the same interests and (b) same ethnic identity but does not share the same interests in activities. Following Nesdale and Lawson (2011), groups had two types of norms: inclusive (e.g., "We invite kids who are different from us") and exclusive (e.g., "We only invite kids who are similar to us").

Two age groups (11- to 12- and 15- to 16-year-olds) were used to test for age-related differences in children's and adolescents' intergroup attitudes and development of group dynamics understanding. Previous developmental intergroup research investigating group norms has used samples ranging from ages 6 to 11 years (Abrams et al., 2009; Nesdale, Maass, et al., 2005), and found that older children become more attuned to group norms than their younger counterparts. It is during adolescence that ethnic group identity becomes most salient (Verkuyten, 2004). In addition, concerns for optimal group functioning and awareness of stereotypic information about other groups become pronounced in adolescence (Killen et al., 2002). Fifteen- and 16-year-olds would also have had ample experience with groups (Brown et al., 1986) and a more complex knowledge about group processes than 11- and 12-year-olds. Therefore, as hypothesized by SRD, a comparison of these two age groups (12- and 16-year-olds) would capture developmental differences in the role of group identity, group norms, and stereotypes in intergroup judgments.

Hypotheses

The hypotheses were derived from the theoretical model guiding the study (SRD; Rutland et al., 2010). First, from an SRD perspective children's inclusive orientations will align with their interest to enhance in-group positivity; thus, it was hypothesized that participants will anticipate their own group (non-Arab American) to be more ethnically inclusive (i.e., include based on shared interest and not ethnic identity) than will the out-group. With age, however, ethnic inclusivity will decrease, given the decline of cross-race friendships (Hallinan & Williams, 1989), increasing examination of one's own ethnic identity, and increasing concerns with group functioning and group dynamics (Killen et al., 2002; Verkuyten, 2004).

Second, as posited by SRD, group norms will impact judgments such that children and adolescents will be more likely to include others irrespective of a person's ethnic identity or shared interest when the group norms were inclusive; however, sensitivity for group norms will increase with age (Abrams et al., 2009; Nesdale & Lawson, 2011; Rutland et al., 2010). Conversely, exclusive group norms will heighten expected ethnic exclusivity by the out-group. Thus, the third hypothesis is regarding differential expectations about one's ethnic in-group versus an ethnic out-group. Based on research evidencing perceptions that out-group dyads are more similar even if they do not share the same interests (McGlothlin & Killen, 2005), it is expected that children will assume the criteria for inclusion (shared interests vs. ethnic identity) will be based on ethnic identity for the ethnic out-group while the criteria for inclusion will be based on shared interest for one's own ethnic in-group. This will be supported by children's and adolescents' reasoning about their inclusion expectations. Therefore, their reasoning will reflect different definitions of "similarities" for their own group in comparison to the Arab American group (e.g., for the Arab American group: "They are more likely to include Hani because he is Arab like them"; for the non-Arab American group; "They will include Omar because he likes the same things").

Finally, based on research showing that stereotypes can be used to justify exclusion (Horn, 2003; Killen et al., 2002), it was expected that participants who make stereotypic associations regarding an Arab out-group will be more inclusive toward an ethnic in-group member than an ethnic out-group member, irrespective of whether they shared the same interests. Given that exclusive group norms decrease favorability toward ethnic out-group members (Nesdale, Maass, et al., 2005), the effects of stereotypes were expected to be more pronounced when groups had exclusive norms.

Method

Participants

Participants included 199 non-Arab American 6th and 10th graders who attended public middle and high schools in a school district serving a low-middle- to middle-income population. The sample consisted of two age groups: 102 children (12-year-olds, $M = 12.08$ years, $SD = 0.49$, 52% female) and 97 adolescents (16-year-olds, $M = 16.16$, $SD = 0.82$, 53% female). The sample included 48% European Americans, 17% African Americans, 8% Asian Americans, 7% Latin American, 15% biracial, and 5% other races or ethnicities. Participants were asked to report their religion and 63% were Christian, 12% were Jewish, 2% were Muslim, and 23% reported other religions and beliefs (e.g., agnostic, atheist, Hindu). Two 6th graders and one 10th grader reported being of Arab descent; these participants were removed from the sample due to the study design, leaving a total of 199 for the final sample.

Procedure

Parental consent forms were distributed with an average of 65% return rate and all students who had parental consent participated (the demographics of the sample revealed the same demographics of the schools). The study was introduced to students as a study about how children think about who to include in or exclude from their group of friends and what things are important to them when making these decisions. On the day of data administration paper-and-pencil questionnaires were distributed to students in groups of 20–25. Students took 30–40 min to complete the questionnaire.

Design

The questionnaire was designed to measure group inclusion judgments based on three factors, each with two levels: (a) group ethnic identity (Arab American and non-Arab American), (b) group norm (exclusive and inclusive), and (c) target of inclusion (ethnic out-group target with similar interests, ethnic in-group target with different interests).

Age (12- and 16-year-olds) and stereotypic associations (stereotypes vs. no stereotypes) were independent variables of interest.

Two versions of the questionnaire were randomly administered (Version 1: 49.7%, $n = 99$; Version 2: 50.3%, $n = 100$, evenly divided by age and gender). Female participants received stories about female characters, and male participants responded to stories about male characters. All participants assessed two hypothetical stories, one about a group of Arab American friends (see Figure 1b) and another about a group of non-Arab American friends who hang out after school (see Figure 1a), each having different group norms. Version 1 included one story about an exclusive Arab American group and another story about an inclusive non-Arab American group; the group norms were reversed in Version 2. For each story, there were two conditions, which required social decisions about inclusion of an: (a) ethnic out-group target with similar interests and (b) ethnic in-group target with different interests.

Participants were first introduced to their own ethnic group (see Figure 1a) and completed a Group Identification Task in which they were told that they belonged to a "non-Arab" American group of friends ("This is your group") in the form of an illustration of four peers with non-Arab names (e.g., Sandra, Michael). The Group Identification Task was modified from Nesdale, Durkin, Maass, and Griffiths (2005) and used to help participants identify with their group. This method has been previously validated in developmental intergroup research, which adopts the minimal group paradigm and novel groups to investigate intergroup attitudes (Dunham, Baron, & Carey, 2011). Participants were asked to give their group a name, choose an end-of-year activity they would like the group to do, and pick a symbol for the group. In previous research, children were found to identify with their groups as measured by their levels of happiness for being part their assigned group after being primed with this task (Nesdale, Durkin, et al., 2005). In the form of three illustrations, participants were shown three types of activities that their group liked to do (e.g., filming, roller blading, trumpet).

Next, on the same page, the Arab American group of friends was introduced ("This is the other group"), through an illustration of four same-gendered "Arab" peers (see Figure 1b). To identify the Arab American group as Arab, each member of the group was given an Arabic name written phonetically using the English alphabet (e.g., Samya, Rami), and was depicted to have darker features such as hair and eyes, but had varying skin tones. The depiction of the Arab American group of friends was intended to represent heterogeneity as Arab Americans can come from numerous regions around the world. The group had similar clothing styles as the American group, so as not to have clothing customs impact children's judgments about each group. A statement written in Arabic, which translates to "Arab group of friends," was shown above the picture of the group. The three activities that the Arab American group liked to do were also indicated through depiction (e.g., painting, tennis, photography). These were similar in genre as the activities of the non-Arab American group as to control for any stereotypic assumptions about the interests of peers from Arab backgrounds. On the next page participants were informed about the norms of each group, which were established as follows.

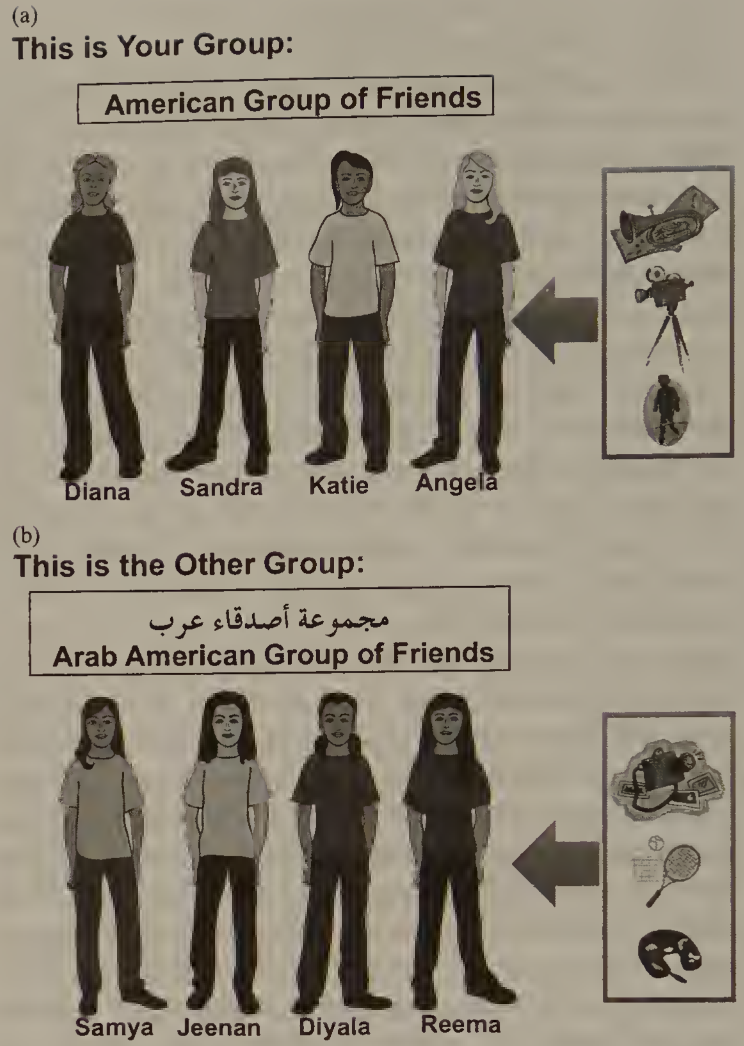


Figure 1. Ethnic peer groups as depicted in the female protocol. (a) Non-Arab American group. (b) Arab American group. © 2011 J.M.K. Tycko, Illustrator.

Exclusive group norm:

> In the past when your/this group of American/Arab American friends, who are your age,

invited others into their group, they would invite only those who were *similar* to them.

Inclusive group norm:

In the past when your/this group of American/Arab American friends, who are your age, invited others into their group, they would invite those who were *different* from them.

Following the description of the norms, participants responded first to a story about their ethnic out-group (Arab American peer group) who had to decide to include a new peer, and responded second to a story about their ethnic in-group. The order of the stories was counterbalanced in a pilot study and showed no order effects so the former order was maintained for both versions of the questionnaire. In each story participants were introduced to an out-group peer with similar interests and an in-group peer with different interests, each seeking entry into the group (e.g., "Hani is Arab American and wants to join the group, he likes these activities"). The activities that each target liked were also depicted pictorially. When the target was an ethnic out-group peer the activities were similar to those depicted for the group, and when the target was an ethnic in-group peer pictures of activities were different from those depicted for the group but also represented varying sports, artistic, and outdoor activities. After being introduced to a target, participants responded to several measures about inclusion and group dynamics regarding the target.

Measures

Inclusion Judgments and Reasoning Assessments

Two assessments were the focus of the current study, (a) group inclusion ("How likely would it be that the group decides to invite Zeina/Julie to join the group?" Likert-type scale: 1 = *really not likely*, to 6 = *really likely*), and (b) group inclusion reasoning ("Why?").

Stereotypes About Arab People Task

This task was administered at the end of the questionnaire and after participants responded to all inclusion judgments and reasoning measures. The task included two items adopted from previous research assessing stereotypes with Jewish and Palestinian children (Bar-Tal, 1996; Cole et al., 2003) but adjusted for use with the current sample based on pilot data: (a) knowledge about Arab culture and (b) stereotypic associations. Two open-ended questions were administered to participants: "Please define who is an Arab" and "What characteristics if any, do you think of when you think of an Arab?" Responses were coded for three types of responses (which included subcategories that were collapsed due to low frequency): (a) *stereotypic associations* included generalizations about physical markers, ethnic customs, religion, negative associations, geography and language (e.g., "They have beards," "They believe in Allah," and "I think of Bin Laden" and "They have an accent"); (b) *factual knowledge* included factual and accurate knowledge about Arab people (e.g., "Someone from the Middle East and speaks Arabic" and "Someone from an Arab country whose parents are Arab"); and (c) *other* included no knowledge or no references to any distinguishing characteristics. Responses were coded, 1 = *full use of the category*, .5 = *partial use*, 0 = *no use of the category*. Partial use of a category occurs when it is cited simultaneously with another category. On the basis of 25% of the interviews ($n = 50$), at least 88% agreement was achieved between three coders, with Cohen's $\kappa = .86$ for interrater reliability.

For descriptive statistics, a new dichotomous Arab knowledge variable was created from the responses to the knowledge about Arab culture item, which divided the sample into those who had factual knowledge about Arab people and those who did not (1 = *factual knowledge*, 0 = *nonfactual knowledge*). To be used both for descriptive statistics and as an independent factor a new dichotomous stereotypes variable was also created to track the presence or absence (1 = *presence*, 0 = *absence*) of stereotypic associations (generalizations about physical markers, ethnic customs, religion, negative associations, geography, and language) in participants' responses to the stereotypic associations item.

Contact With Individuals of Arab Descent Measure

Participants' level of contact with peers of Arab descent was measured for descriptive purposes, using the Contact with Individuals of Arab Descent assessment. This included five items reflecting varying levels of contact (e.g., "How many kids in your neighborhood are Arab?" Likert-type scale: 1 = *none*, 2 = *a few*, 3 = *half*, 4 = *most*, and 5 = all, or "How often do you hang out with kids who are Arab?" Likert-type scale: 1 = *never*, 2 = *a little*, 3 = *sometimes*, 4 = *often*, 5 = *always*). Factor analysis was conducted on all five items, showing that they load on one factor, representing participants' level of contact with Arab peers and explaining 50.1% of variance in all five variables. Contact with individuals of Arab descent was low for this sample, $M = 1.88$, and $SD = 0.58$ (for the Likert-type scale data). In addition, participants were asked about the ethnic composition of their friendship group. This item was worded as the following: ethnic or racial composition of friendships ("How many of your friends are the same ethnicity or race as you?" Likert-type scale: 1 = *none*, 5 = *all*). Forty percent of participants reported having more than half their friends of a similar race, while 60% had less than half their friends of the same race or ethnicity.

Reasoning and Coding Reliability

Participants' justifications were coded using a coding system composed of categories drawn from social domain theory (Smetana, 2006), as well as based on the results of pilot testing. Analyses were conducted using the three most frequently used justifications, which were categories used with a proportion of .10 or higher. For this study these codes included: (a) group functioning (e.g., "He won't fit in" or "It doesn't affect the group"), (b) activity preferences (e.g., "She likes to do different activities" or "He likes tennis just like them"), and (c) ethnic identity (e.g., "He's American which is different from them" or "She's Arab American, they would like her"). Justification responses for inclusion judgments could have a maximum of two codes; thus, responses were coded as 1 = *full use of the category*, .5 = *partial use*, 0 = *no use of the category*. On the basis of 25% of the interviews ($n = 50$), at least 89% agreement was achieved between three coders with Cohen's $\kappa = .88$ for interrater reliability.

Plan for Analyses

Hypotheses regarding group inclusion judgments were tested using one-sample t tests and repeated measures analysis of variance (ANOVA). Planned comparisons were conducted to test expected differences between factors, and Bonferroni comparisons were conducted as follow-up tests on the ANOVA to control for Type I errors. The primary variables of interest for this study were group ethnic identity (Arab American, non-Arab American), group norm (inclusive, exclusive), the target of inclusion (ethnic out-group with similar interests, ethnic in-group with different interests), age, and stereotypic associations as independent factors. Although no gender differences were hypothesized and this study was not designed to address questions about gender, gender was maintained in the analyses, given that past research has shown differential social exclusion judgments based on gender (Killen et al., 2002) due to social status differences (e.g., status based on gender). Dependent measures in these analyses were group inclusion judgments. Homogeneity of covariance and variance were examined and showed nonsignificant Box's statistical test and Levene's statistical test, respectively. Differences between the proportions of the top three types of reasoning used for the group inclusion reasoning measure were assessed using repeated measures ANOVAs. ANOVAs were used to analyze proportions due to repeated measures designs, which are not appropriate for logistic regressions (see footnote 4 in Wainryb, Shaw, Laupa, & Smith, 2001). Descriptive analyses using independent samples t tests were conducted on the stereotypes about Arab people task to assess group differences based on age, gender, religion, and ethnic composition of friendships (i.e., more than half of friends of the same ethnicity and less than half of friends the same ethnicity).

Results

Judgments and Reasoning About Group Inclusion

Group Inclusion Judgments

Will the group include X? Initially, one-sample t tests were conducted to assess whether participants' group inclusion judgments in each target condition in each of the four stories (stories: exclusive Arab American group, exclusive non-Arab American group, inclusive Arab American group, inclusive non-Arab American group; see Figure 2 for all the conditions), differed from a 3.5 midpoint, which indicated an average 50% likelihood of including the target in the group. Participants expected groups to be inclusive with more than 50% likelihood of inclusion in all conditions at $p < .001$, except for two. These two conditions were when an exclusive Arab American group was deciding to include an ethnic out-group target with similar interests, and when an exclusive non-Arab American group was deciding to include an ethnic in-group target with different interests (see Figure 2 for test statistics).

Differences between ethnic groups. The first hypothesis regarding whether participants would anticipate their own group to be more ethnically inclusive than the out-group was examined with the group inclusion measures as dependent variables in two separate 2 (age group: 12 and

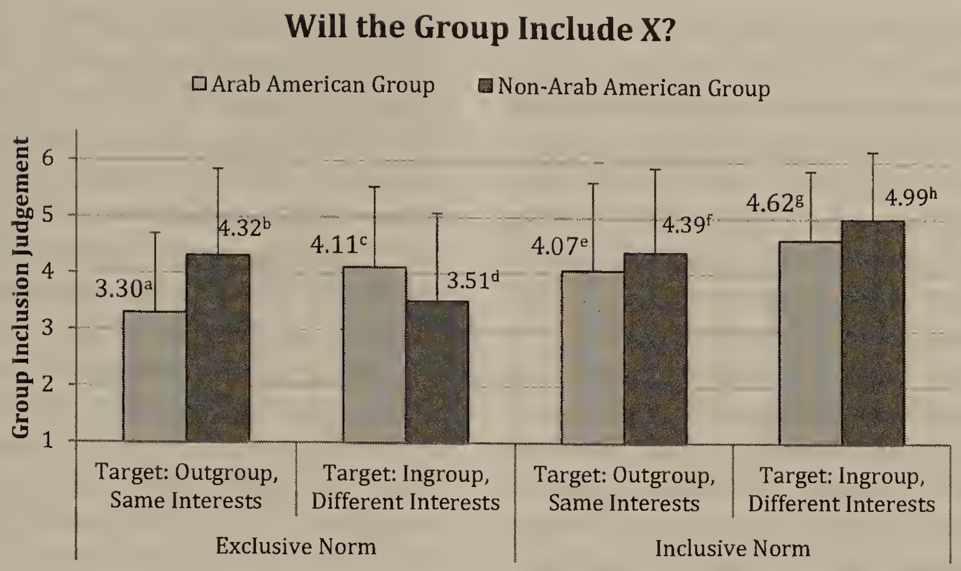


Figure 2. Group inclusion judgments for both targets by ethnic group identity and group norm. Inclusion judgments were made on a Likert-type scale: 1 = *really not likely to* 6 = *really likely*. Error bars represent standard deviations.
[a]*ns* compared to 3.5 midpoint inclusion judgment. [b]$t(99) = 5.47$, $p < .001$, Cohen's $d = 0.55$. [c]$t(97) = 4.19$, $p < .001$, Cohen's $d = 0.42$. [d]*ns* compared to 3.5 midpoint inclusion judgment. [e]$t(97) = 3.60$, $p < .001$, Cohen's $d = 0.36$. [f]$t(98) = 5.57$, $p < .001$, Cohen's $d = 0.56$. [g]$t(99) = 9.20$, $p < .001$, Cohen's $d = 0.92$. [h]$t(97) = 11.31$, $p < .001$, Cohen's $d = 1.00$.

16 years) × 2 (group norm: inclusive, exclusive) × 2 (gender: female, male) × 2 (group ethnic identity: Arab American, non-Arab American) ANOVAs with repeated measures on the last factor. One ANOVA was for assessing inclusion of an ethnic out-group target with similar interests and the other was for an ethnic in-group target with different interests. Ethnic inclusivity was demonstrated by high ratings for the inclusion of an ethnic out-group target with similar interests (i.e., shared interests trumps ethnic identity). Ethnic exclusivity on the other hand was demonstrated by high ratings for inclusion of an ethnic in-group target who did not share the same interests as the group (i.e., ethnic identity trumps shared interests). Group norms and age were included in the repeated measures ANOVAs to examine hypotheses related to exclusive and inclusive group norms as well as a decrease in ethnic inclusivity in adolescents.

Ethnic out-group target with similar interests. Expectations were confirmed for the ethnic out-group target. First, a main effect for group ethnic identity was captured, $F(1, 188) = 14.64$, $p < .001$, $\eta_p^2 = .07$. Findings showed that participants perceived their own group to be more inclusive toward an Arab American (out-group target) with shared interests ($M = 4.36$, $SD = 1.52$) than an Arab American group would be toward a non-Arab American with similar interests as the group ($M = 3.68$, $SD = 1.53$). Thus, they expected their own group to demonstrate ethnic inclusivity more than the Arab American out-group.

Second, a Group Ethnic Identity × Group Norm interaction effect was found, $F(1, 188) = 6.19$, $p < .05$, $\eta_p^2 = .03$, confirming expectations about the effects of group norms, indicating the main effect was driven by both groups having exclusive norms ("We like kids who are similar"). Thus, in the exclusive norm stories, the Arab American group was viewed to be less inclusive than the non-Arab American group toward someone with a different ethnic identity (shared interests), $p < .001$ (see Figure 2). When the groups had inclusive norms ("We like kids who are different"), however, no differences were found for including an ethnic out-group target. Third, a main effect for age, $F(1, 188) = 4.08$, $p < .05$, $\eta_p^2 = .02$, was found, showing that older participants perceived both ethnic groups to be less inclusive toward an out-group target than younger participants (12-year-olds: $M = 4.16$, $SD = 0.87$; 16-year-olds: $M = 3.89$, $SD = 0.87$). This confirmed our hypotheses that ethnic inclusivity will decrease with age.

Ethnic in-group target with different interests. When judging inclusivity toward an ethnic in-group target with different interests, no main effect was found based on the ethnic identity of the group; both groups were viewed to be inclusive toward an ethnic in-group target with different interests (Non-Arab American group: $M = 4.24$, $SD = 1.58$; Arab American group: $M = 4.37$, $SD = 1.35$). A Group Ethnic Identity × Group Norm interaction effect was, however, found, $F(1, 188) = 51.97$, $p < .001$, $\eta_p^2 = .21$. In this case, when the groups had an exclusive norm, one's own non-Arab American group was viewed to be less inclusive than the Arab American group toward someone with different interests in activities (same ethnicity), $p < .01$ (see Figure 2). When groups had inclusive norms, however, the finding was reversed. Participants expected their group to be more inclusive than the Arab American group, $p < .05$. No age-related effects were found for inclusion of an in-group target with different interests.

In summary, when judging the likelihood of including an ethnic out-group who shares the same interests as the group, participants expected their own group to be more ethnically inclusive than an Arab American group. This was the case, though, only when both groups had exclusive norms. Irrespective of the ethnic identity of the group, ethnic inclusivity decreased with age. No age effects or overall ethnic group differences were found for inclusivity of an ethnic in-group target who has different interests. Both inclusive and exclusive group norms impacted how children made inclusion judgments regarding an ethnic in-group target (different interests) in their own group, but norms had no impact for inclusion of the same target in an Arab American group.

Shared Interests Versus Ethnic Identity

To examine whether participants expected inclusion criteria to differ for each group, inclusion judgment ratings for an ethnic in-group target were compared against ratings for an ethnic out-group target for each peer group separately. Therefore, two 2 (age group: 12 years, 16 years) × 2 (group norm: inclusive, exclusive) × 2 (gender: female, male) × 2 (target of inclusion: out-group member with similar interests, in-group member with different interests) ANOVAs with repeated measures on the last factor were conducted with group inclusion judgments for each group as the dependent variable. The hypotheses regarding differences in judgments based on group norms and age were also examined by the inclusion of these variables as factors in the repeated measures ANOVAs.

Non-Arab American group. First when examining the results for the non-Arab American group there was no main effect for the target of inclusion. This indicated that participants expected their own group to be equally likely to include both targets. Second, confirming predictions about the role of group norms, a main effect for group norm was found showing participants' overall inclusion judgments regarding both targets were lower when their group had an exclusive norm than when it had an inclusive norm, $F(1, 189) = 33.43$, $p < .001$, $\eta_p^2 = .15$ (exclusive norm: $M = 3.91$, $SD = 0.91$; inclusive norm: $M = 4.66$, $SD = 0.91$). However, these main effects were qualified by a Target of Inclusion × Group Norm interaction effect, $F(1, 189) = 18.27$, $p < .001$, $\eta_p^2 = .08$. As shown in Figure 2, the interaction indicated confirmation of the hypothesis that shared interests would trump ethnic identity for the non-Arab American group when the group had an exclusive norm. Thus, participants expected their own exclusive group would be more likely to include an ethnic out-group target who shared the same interests as the group than it would be to include an ethnic in-group target who did not share the same interests, $p = .001$. Although participants were likely to include both targets when their group had an inclusive norm ("We like kids who are different"), they expected their group would more likely include a target who liked different activities, $p < .05$, thus reaffirming their focus on activities

In addition, participants were especially sensitive to their group's norm when making decisions about including an ethnic in-group target (different interests) but not when making decisions about an ethnic out-group target (shared interests). Thus, in the case of an ethnic in-group peer, participants expected their group to be less inclusive when it had an exclusive norm than when it had an inclusive norm, $p < .001$ (exclusive norm: $M = 3.51$, $SD = 1.56$; inclusive norm: $M = 4.95$, $SD = 1.27$). Irrespective of one's own group norm, participants expected that their group would be inclusive toward an ethnic out-group target with similar interests (exclusive norm: $M = 4.32$, $SD = 1.53$; inclusive norm: $M = 4.37$, $SD = 1.54$). Contrary to expectations that adolescents would be ethnically less inclusive than 12-year-olds when groups had exclusive norms, there was no Age × Group Norm effect.

Arab American group. Different patterns emerged when anticipating the inclusivity of an Arab American group. First, unlike evaluations of their own group when evaluating an Arab American, a main effect for target of inclusion, $F(1, 187) = 18.38$, $p < .001$, $\eta_p^2 = .09$, was found showing that participants expected the Arab American target group would be less likely to include a non-Arab American target than an ethnic in-group target (ethnic out-group, similar interests: $M = 3.67$, $SD = 1.53$; ethnic in-group, different interests: $M = 4.36$, $SD = 1.33$). Second, a main effect for group norm was also found, indicating that, irrespective of the target's identity, inclusion judgments matched the group's norm, $F(1, 187) = 30.97$, $p < .001$, $\eta_p^2 = .14$

(exclusive norm: $M = 3.68$, $SD = 0.86$; inclusive norm: $M = 4.36$; $SD = 0.86$). No interaction effect was found for Target of Inclusion × Group Norm. Therefore, the criteria for inclusion did not change as a function of the group's norm.

In summary, shared interests took priority as the criterion for inclusion over ethnic identity when both 12- and 16-year-olds anticipated their own group's inclusivity. However, expectations for an Arab American group were different, showing that participants expected this group to give priority to ethnic identity over shared interests, irrespective of the group norm. Inclusive group norms generally elicited more inclusive judgments for both groups.

Reasoning About Group Inclusion

To understand how participants interpreted "similarity" for each ethnic group, participants' reasoning about group inclusion judgments under the exclusive group norm stories (e.g., "We invite kids who are similar") was analyzed. The three most frequently used justifications were group functioning, activities preferences, and ethnic identity. Two separate 2 (age group: 12 and 16 years) × 2 (gender: female, male) × 2 (group ethnic identity: Arab American, non-Arab American) × 3 (reasoning: group functioning, activity preferences, ethnic identity) ANOVAs with repeated measures on the last factor were conducted using proportion of reasoning for inclusion of an ethnic out-group target (first ANOVA) and inclusion of an ethnic in-group target (second ANOVA).

Ethnic out-group target with similar interests. When judging the inclusion likelihood of an ethnic out-group target with similar interests in a group with an exclusive norm, first, a main effect for the type of reasoning used was found, $F(2, 378) = 4.76$, $p < .05$, $\eta_p^2 = .02$. This finding shows that participants reasoned about group decisions regarding inclusion of this target by referencing activity preferences more so than ethnic identity, $p < .01$, and approximately similar proportions of group functioning (see Table 1). As shown in Table 1, a Reasoning × Group Ethnic Identity interaction effect was found, $F(2, 378) = 4.05$, $p < .05$, $\eta_p^2 = .02$, showing that participants referenced ethnic identity more when judging likelihood of inclusion by the Arab American group than they did for inclusion by their own non-Arab American group, $p < .01$.

In addition, only when judging their own group's decision to include an out-group target did participants use more group functioning and activity preference reasoning than ethnic identity, $ps < .001$ (see Table 1). Thus, when evaluating a group with an exclusive group norm, participants attended more to ethnic identity ("She's not Arab") when judging inclusion of an ethnic out-group target with similar interests in the Arab American group and attended more to group functioning ("He fits in better") and activity preferences ("They like to do the same things") when thinking about their own group.

Ethnic in-group target with different interests. When the group had an exclusive norm, reasoning about decisions to include an ethnic in-group target with different interests also differed based on the ethnic identity of the group. Overall, participants used similar proportions of reasoning except this changed based on the ethnic group they were judging. A Reasoning × Group Ethnic Identity interaction effect was found, $F(2, 374) = 8.89$, $p < .001$, $\eta_p^2 = .04$. This effect showed that participants reasoned based on group functioning and activity preferences more so than ethnic identity when judging inclusion of an in-group target into their own group, both $ps < .01$ (see Table 1). Significantly more references for group functioning, $p < .01$, but less references to ethnic identity, $p < .001$, were made when reasoning about inclusion into a non-Arab American group compared to reasoning about inclusion into an Arab American group (see Table 1). They referenced ethnic identity more than group functioning when reasoning about inclusion of an Arab American target in the Arab

Table 1
Proportion of Reasoning Used to Judge Inclusion of Both Targets in an Exclusive Group

Condition	Reasoning	Arab American group *M (SD)*	Non-Arab American group *M (SD)*	Total *M (SD)*
Target: Out-group, similar interests	Group functioning	0.28 (0.41)	0.37 (0.47)	0.32 (0.44)
	Activity preferences	0.32 (0.37)	0.40 (0.45)	0.36 (0.41)
	Ethnic identity	0.29 (.39)	0.14 (0.30)	0.21 (0.35)
Target: In-group, different interests	Group functioning	0.20 (0.37)	0.37 (0.47)	0.29 (0.43)
	Activity preferences	0.30 (0.38)	0.34 (0.45)	0.32 (0.41)
	Ethnic identity	0.39 (0.44)	0.15 (0.32)	0.27 (0.40)

Note. $N = 199$.

American group, $p < .01$. In sum, when evaluating inclusion of an ethnic in-group target in their own exclusive group, participants were concerned more with group functioning and interests in activities ("She likes different things"). When making similar judgments regarding an exclusive Arab American group, they frequently referenced ethnic identity ("He's Arab American, like they are"). Overall, the findings for reasoning about both targets indicated that participants gave more priority to individual level characteristics such as shared interests in activities when thinking about their own ethnic group but expected an ethnic out-group to be more concerned with ethnic identity.

The Role of Stereotypes

Stereotypes About Arab People Task

Before analyzing the effects of stereotypes on participants' group inclusion judgments, descriptive analyses were conducted on responses to the stereotypes about Arab people task. First, the Arab knowledge variable was used as the dependent variable in independent sample *t*-test analyses to assess participants' group differences in their knowledge about Arab people. This dichotomous variable indicated that 33% of participants reported having factual knowledge about Arabs (*Arab knowledge* = 1) and the remaining 67% did not (*Arab knowledge* = 0). A large effect size for age was found, showing older participants had more knowledge about Arab people than younger participants, $t(168) = -5.89$, $p < .001$, Cohen's $d = -0.86$ (12-year-olds: $M = 0.15$, $SD = 0.36$; 16-year-olds: $M = 0.52$, $SD = 0.50$). Knowledge did not vary by gender, religion, or ethnic composition of friendships.

When analyzing differences in the proportion of generalizations made about Arab people by participants, the stereotypes variable was used as the dependent variable in independent sample *t* tests. Forty-seven percent ($n = 88$) of participants did not report stereotypic associations with Arab people, while 53% did ($n = 99$). Older participants reported more stereotypic associations than younger participants, revealing a moderate effect size, $t(185) = -4.90$, $p < .001$, Cohen's $d = -0.73$ (12-year-olds: $M = 0.36$, $SD = 0.48$; 16-year-olds: $M = 0.70$, $SD = 0.46$). Stereotypic associations did not vary by gender, religion, or ethnic composition of friendships. Participants' stereotypic associations did not significantly differ based on their knowledge of Arab people.

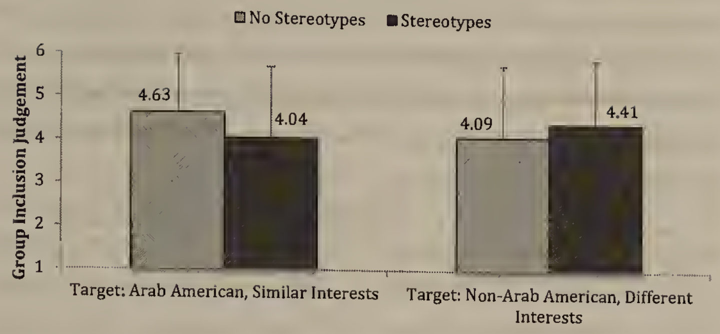
Effects of Stereotypes on Group Inclusion Judgments

To test hypotheses related to stereotypes and judgments about group inclusion the stereotypes variable was used as a fixed factor in repeated measures ANOVAs with age and group norm. No gender effects were investigated because there were no hypotheses pertaining to gender and stereotypic associations. Additionally, the proportion of stereotypic associations made did not differ by gender. Two 2 (age group: 12 and 16 years) × 2 (group norm: inclusive, exclusive) × 2 (stereotypes: presence, absence) × 2 (target of inclusion: out-group member with similar interests, in-group member with different interests) ANOVAs with repeated measures on the last factor were conducted on group inclusion judgments. One analysis was conducted for judgments about their own group and another for judgments about an Arab American group. Only effects related to stereotypic associations will be reported as other effects have already been reported above.

When making inclusion judgments about their own ethnic group, as expected, a Target of Inclusion × Stereotypes interaction effect was found, $F(1, 178) = 6.79$, $p < .01$, $\eta_p^2 = .03$. As depicted in Figure 3a, this finding shows that those who made stereotypic associations about Arabs were less inclined to think their own group would include an Arab American than those who did not make stereotypic associations, $p < .05$. In addition, those who did not make stereotypic associations were more inclusive toward an Arab American who shared the same interest in activities than toward a non-Arab American target who did not share the same interests as the group, $p < .05$ (see Figure 3a). When judging inclusivity of the Arab American group toward either target, no differences were found between those who made stereotypic association and those who did not. No main effects for stereotypes were found in either analysis.

Planned pairwise comparisons were conducted to test hypotheses pertaining to the role of group norms and how they may influence those who hold stereotypes. These analyses indicated that stereotypes proved to play a statistically significant role for participants judging inclusion of an Arab American, $F(1, 178) = 6.81$, $p = .01$, $\eta_p^2 = .03$, when one's own group had an exclusive norm but not when the group had an inclusive norm (see Figure 3b). These findings indicate that participants who held stereotypic knowledge about Arab people expected their group to be less inclusive toward an Arab American peer, especially when their group had an

(a) Non-Arab American Group



(b) Non-Arab American Group
Target: Arab American with Similar Interests

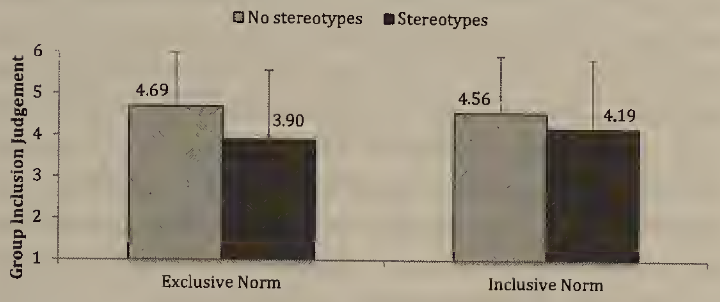


Figure 3. (a) Role of stereotypic associations on the non-Arab American group's inclusion judgments about ethnic out-group and in-group targets. (b) Group inclusion: the role of group norms on those who make stereotypic associations. Inclusion judgments were made on a Likert-type scale: 1 = *really not likely* to 6 = *really likely*. Error bars represent standard deviations. *N* = 199.

exclusive norm. However, when their group had an inclusive norm those with stereotypes did not differ in their inclusion expectations from those without stereotypes. No stereotype by group norm effects were found for judgments about the Arab American group.

Discussion

The novel findings of this study revealed that non-Arab American children and adolescents expected that an ethnic out-group (of Arab descent) would value ethnic identity over individual characteristics, such as shared interests in activities, when given the opportunity to invite a new person to join their group. Conversely they anticipated that their in-group would be inclusive, valuing shared interests in activities more than ethnic identity. In addition, stereotypic associations were found to be directly related to children's and adolescents' inclusion expectations. These findings are important and generalize to other cultural contexts. Many places around the globe are experiencing new immigrant groups. In these contexts little prior intergroup contact exists but negative stereotypes are associated with the newly arrived groups. One implication of these findings is that a contributing factor to pervasive segregation of majority groups with newly immigrated minority groups is the expectation by the majority group that the new group "prefers to be with their own kind."

A limitation with this expectation is that it assumes homogeneity of the out-group (with little recognition of the intragroup variation regarding choice of friends and companions) and a generalized view that overtures of inclusion are not warranted. Yet, individuals in newly immigrated groups need to both socialize with their own groups and integrate with the majority group for healthy adjustment within their host communities (Phinney, 2008; Verkuyten, 2004). Moreover, when assumptions of exclusivity are associated with stereotypes, the assumption may not be founded on actual information about preferences for exclusivity but expectations due to unfamiliarity with the out-group or even due to essentialist beliefs that out-groups have inherently different values and live by different rules than one's in-group.

Children and adolescents gave priority to shared interests in activities over ethnicity as the criteria for inclusion by their own group. This is consistent with research on young children's perceptions of similarity between cross-race dyads that share the same interests (McGlothlin & Killen, 2005). Past research on dyadic friendships in younger participants has shown that variables such as language (Kinzler, Shutts, Dejesus, & Spelke, 2009) also trump ethnicity in social preferences of young children (5–6 years).

The current study extended these previous findings in several ways. First, in contrast to social preference or perceptions of similarity, the current findings were demonstrated using a social-contextual method, in which judgments and reasoning were analyzed for inclusion decisions by both the in-group and the out-group as well as for when the groups had inclusive or exclusive norms. Second, as predicted from an SRD perspective, the finding that shared interests trumped ethnicity was context specific, such that shared interests were expected to matter more for one's own group than for an ethnic out-group. Third, the current findings were shown with an older sample (12- and 16-year-olds) than previously studied. This is important as adolescence is a developmental period in which group identity is highly salient, and group norms guide decisions about peer relationships. Fourth, having an inclusive in-group norm ("We like kids who are different") allowed participants to equally consider both variables (i.e., ethnicity and shared interests) when making inclusion judgments. Finally, while previous research shows that shared interests trumped race regarding expectations about one's own group, in the current study ethnicity trumped shared interests for expectations about an ethnic out-group.

In this study, differential expectations about inclusion based on ethnic group identity were robust as they appeared in both children's judgments and reasoning. Participants' justifications for their inclusion judgments when both groups had exclusive norms focused on ethnic identity (e.g., "She's Arab like them") when reasoning about the Arab American group and on group functioning (e.g., "They'll get along better") or activity preferences (e.g., "He likes the same things as the group") when reasoning about their own non-Arab American group. These findings extend SRD by showing that both children and adolescents use different criteria for group dynamics when groups have different ethnic identities.

Additionally, the findings support the view that participants' judgments about the Arab American group are a form of implicit stereotyping, reflecting intergroup attribution error, whereby a misattribution is made based on a social category (Hewstone, 1990). In this case the misattribution being made is that an Arab American group would not be inclusive toward a non-Arab peer; therefore, when seeking entry into an ethnic out-group (Arab American peer group), "being similar" was interpreted as being of a same ethnicity, as opposed to sharing the same interests. Given that "similar" was never defined within the questionnaire, children's differential interpretations about what "similar" means can also reflect an indirect, potentially implicit form of bias that favors their in-group (i.e., interpreting similarity based on shared interests as opposed to ethnicity). To further confirm such interpretations, future research should also examine children's own individual inclusive preferences and compare them to their expectations about peer groups' inclusivity.

Given research showing that children are less likely to want to be friends with those they perceived as having exclusive norms (Tropp et al., 2014), the overgeneralization (or misattribution) found in the current study is likely to hinder the development of cross-ethnic friendships between Arab American and non-Arab American youth. Perceiving one's peers to be exclusive solely on the basis of their group membership assumes homogeneity of the group, which undermines an individual's own choice about potential friendships. This misattribution was found in both 12- and 16-year-olds, which provides a potential explanation as to why cross-ethnic friendships decline in adolescence, and also suggests that the phenomenon begins in early adolescence for friendships with peers of Arab descent.

Age differences appeared when judging the likelihood that a group would include an ethnic out-group target. Consistent with expectations from SRD, these findings showed that 16-year-olds found ethnic identity to be more salient than 12-year-olds because they expected both groups to be less inclusive toward ethnic out-group targets despite their sharing an interest in activities. Past research shows that although both children and adolescents reject exclusion, children (9- to 11-year-olds) reject it more than adolescents (Killen et al., 2002). These previous findings are qualified by adolescents' concerns with optimizing group dynamics, and the perception that diversity in ethnic identity can hinder group functioning (Rutland et al., 2010).

It was expected that adolescents would be more attuned to group norms than 12-year-olds, given their more extensive experience with groups (Brown et al., 1986). However, no age differences were found based on the type of group norm; both 12- and 16-year-olds responded in similar ways when the group had either an exclusive or inclusive norm. This may be because by 12 years children have had experience navigating through exclusive and inclusive group behavior, and have established a good understanding of group norms that fall within the domain of exclusion and inclusion. Twelve-year-olds, however, may not have had experience with other types of group norms, such as those related to experiences in adolescence (e.g., student council, financial scholarship, risky behavior), and therefore differences between preadolescents and adolescents may be more pronounced in those contexts as opposed to the norms presented in the current study. What was clear though, and consistent with past research (Nesdale, Maass, et al., 2005), was that defining a norm in terms of "liking others who are similar" resulted in less inclusive judgments and defining a norm in terms of "liking others who are different" elicited more inclusive judgments.

Participants' sensitivity to group norms was more nuanced when making inclusion judgments about their own group. They expected that group norms would influence inclusion of an ethnic in-group peer who did not share the same interests in activities in their own group. Inclusion of an ethnic out-group member who shared the same interests as the group, however, was not affected by their group's norm. This nuanced effect of group norms did not manifest when making judgments about inclusivity of an Arab American group. These findings extend SRD theory (Rutland et al., 2010) by showing that group norms are perceived to play a different role for ethnic out-groups than they are for one's own ethnic group. In an effort to fully grasp the role of group norms in intergroup relationships, further research is needed to investigate how other out-group norms are interpreted in comparison to one's own group norms (e.g., freedom of expression).

A central goal of this study was to assess the effects of stereotypes on children's and adolescents' inclusion judgment through the direct measurement of stereotypic knowledge. It was found that children and adolescents who readily made stereotypic associations perceived their own group to be less inclusive toward ethnic out-group targets despite their sharing the same interest in activities as the groups. Those who did not make stereotypic associations (47%) were more inclusive toward ethnic out-group targets. Importantly, the effects of stereotypic associations on inclusion decisions were strongest when the group had an exclusive norm and nonexistent when it had an inclusive norm. Inclusive group norms lessened the effects of stereotypes, thus highlighting the significance of fostering ethnic inclusivity in classrooms and schools, as theorized by Nesdale and Lawson (2011). While Nesdale and colleagues have shown the effects for group norms on how favorable children are toward out-group members, the current study extended these findings to inclusion judgments of both children and adolescents. This finding has important implications for stereotype and prejudice reduction interventions. It implies that inclusive norms can mitigate the effects of stereotypes on children's inclusion decisions.

There were no age-related differences in how stereotypes affected inclusion judgments. However, in contrast to only one-third of the 12-year-olds, the majority (70%) of the 16-year-olds reported stereotypic associations. Moreover, 16-year-olds were less inclusive toward out-group targets than were the 12-year-olds, perhaps because the stereotypic associations contributed to their less inclusive orientations; however, there is only indirect evidence for this given that minimal references to stereotypes (< 10%) were explicitly made in adolescents' reasoning about their judgments. Research on social identity development shows that in late childhood through adolescence, concepts about group identities are based on beliefs, values, status differences, and stereotypes (Bennett, 2011). While the current study sheds light on these age-related shifts, a direct connection between stereotypic expectations and exclusion judgments in adolescents remains to be further investigated. In addition, the level of identification with one's own group is expected to increase in adolescence, and the changing demographics of the United States suggests that ethnic minority children make up more than half of the youth population (Cabrera, 2013); therefore, future studies should examine the level of identification with a superordinate American identity and its impact on inclusion judgments of both ethnic majority and minority children.

Conclusions and Future Directions

Understanding attitudes about out-group peers is important given the increasing global context of child development. Children are growing up in cultures and communities with new immigrants from around the world. The findings in this study provide teachers, counselors, and administrators with some guidance on how to best facilitate positive interethnic relationships. The effects of stereotypic associations and the tendency to misattribute that an ethnic out-group would be ethnically exclusive can be directly addressed by the promotion of inclusive peer group norms. Therefore, schools, classrooms, or groups that foster inclusivity by embracing differences (e.g., ethnic and those based on interests in activities) and ethnic openness can combat the influence of stereotypes on children's and adolescents' social decision making in exclusion contexts. This finding supports research assessing the effects of multiethnic education on children's intergroup favorability evaluations (Verkuyten, 2008).

While the target ethnic out-group in this study was composed of peers of Arab descent, the current findings may be generalized to contexts in which newly immigrated groups associated with negative stereotypes encounter rejection and segregation. Around the world, communities are changing as groups that differ based on ethnicity, culture, and religion (and other group membership categories) are living together for the first time. How children and adolescents conceptualize out-groups with low contact (and often high levels of stereotyping) has to be investigated to create effective interventions. Future research could explore questions about group-based exclusion versus interest-based exclusion with other group categories (as well as other ethnic groups).

Importantly, future research could more specifically examine how different ethnic minority groups perceive other groups in terms of stereotypes and inclusivity. In this study, we focused on two groups because we did not find any significant differences for the different ethnic groups in the "non-Arab American" category. However, more detailed assessments of stereotypes as well as group identity may reveal new relationships and forms of bias. Targeting samples of different ethnic groups in the United States and measuring their level of identification with their ethnic group as well as their identification as a broader American identity would contribute to understanding group categories that enhance inclusive orientations in different contexts.

Exploring the role of stereotypes on inclusive judgments with a sample that has varying levels of contact can inform whether contact will also help mitigate the effects of stereotypes on intergroup exclusion. As indicated by the current findings, a non-Arab perspective on intergroup exclusion was informative. However, it is imperative to conduct research with both ethnic majority and minority groups as both perspectives are essential for understanding group dynamics. In the case of the present study, obtaining a sample with a large representation of Arab participants (for statistical analyses) was not feasible. Having both perspectives, however, provides valuable information for understanding whether minority groups expect majority groups to focus on culture over shared interests. In addition, a one-dimensional stereotype assessment was used in this study, mainly because little was known about the content of the stereotypic associations children made with people of Arab descent. While negative stereotypes were cited, most stereotypic associations in the study represented neutral generalizations such as those related to physical appearance, language, and religious beliefs. Further research must be carried out with different and larger samples to capture specific effects of stereotypes that vary in content.

Overall, the findings from the current study reveal the complex ways in which children and adolescents reason about intergroup exclusion encounters, providing further insight into the development of intergroup attitudes. This information can be used to inform intervention-based research aimed at reducing prejudicial attitudes in social relationships through the promotion of inclusive group norms that combat children's and adolescents' stereotypes and misattributions about ethnic out-groups.

References

Abrams, D., Rutland, A., Pelletier, J., & Ferrell, J. M. (2009). Children's group nous: Understanding and applying peer exclusion within and between groups. *Child Development, 80,* 224–243. doi:10.1111/j.1467-8624.2008.01256.x

Bar-Tal, D. (1996). Development of social categories and stereotypes in early childhood: The case of "the Arab" concept formation, stereotype and attitudes by Jewish children in Israel. *International Journal of Intercultural Relations, 20,* 341–370. doi:10.1016/0147-1767(96)00023-5

Bennett, M. (2011). Children's social identities. *Infant & Child Development, 20,* 353–363. doi:10.1002/icd.741

Brown, B., Eicher, S., & Petrie, S. (1986). The importance of peer group ("crowd") affiliation in adolescence. *Journal of Adolescence, 9,* 73–96. doi:10.1016/S0140-1971(86)80029-X

Cabrera, N. J. (2013). Positive development of minority children. *Social Policy Report, 27,* 1–22.

Cole, C., Arafat, C., Tidhar, C., Zidan, W. T., Fox, N. A., Killen, M., . . . Yung, F. (2003). The educational impact of Rechov Sumsum/Shara'a Simsim, a television series for Israeli and Palestinian children. *International Journal of Behavioral Development, 27,* 409–422. doi:10.1080/01650250344000019

DeRosier, M. (2004). The social climate of schools in the aftermath of 9–11: Ethnic differences in children's peer relations. *Journal of School Violence, 3,* 5–16. doi:10.1300/J202v03n01_02

Dunham, Y., Baron, A. S., & Carey, S. (2011). Consequences of "minimal" group affiliations in children. *Child Development, 82,* 793–811. doi:10.1111/j.1467-8624.2011.01577.x

Flanagan, C., Syvertsen, A., Gill, S., Gallay, L., & Cumsille, P. (2009). Ethnic awareness, prejudice, and civic commitments in four ethnic groups of American adolescents. *Journal of Youth and Adolescence, 38,* 500–518. doi:10.1007/s10964-009-9394-z

Hallinan, M., & Williams, R. (1989). Interracial friendship choices in secondary schools. *American Sociological Review, 54,* 67–78. doi:10.2307/2095662

Hewstone, M. (1990). The "ultimate attribution error?" A review of the literature on intergroup causal attribution. *European Journal of Social Psychology, 20,* 311–335. doi:10.1002/ejsp.2420200404

Hitti, A., Mulvey, K. L., Rutland, A., Abrams, D., & Killen, M. (2014). When is it okay to exclude a member of the ingroup? Children's and adolescents' social reasoning. *Social Development, 23,* 451–469. doi:10.1111/sode.12047

Horn, S. (2003). Adolescents' reasoning about exclusion from social groups. *Developmental Psychology, 39,* 71–84. doi:10.1037/0012-1649.39.1.71

Ibish, H. (2008). *Report of hate crimes and discrimination against Arab Americans: 2003–2007.* Washington, DC: Arab American Anti-Discrimination Committee.

Killen, M., Lee-Kim, J., McGlothlin, H., & Stangor, C. (2002). How children and adolescents evaluate gender and racial exclusion. *Monographs for the Society for Research in Child Development, 67*(4, Serial No. 271). doi:10.1111/1540-5834.00218

Killen, M., Mulvey, K. L., & Hitti, A. (2013). Social exclusion: A developmental intergroup perspective. *Child Development, 84,* 772–790. doi:10.1111/cdev.12012

Killen, M., & Rutland, A. (2011). *Children and social exclusion: Morality, prejudice, and group identity*. New York, NY: Wiley/Blackwell.

Kinzler, K., Shutts, K., Dejesus, J., & Spelke, E. (2009). Accent trumps race in guiding children's social preferences. *Social Cognition, 27,* 623–634. doi:10.1521/soco.2009.27.4.623

Mähönen, T., Jasinskaja-Lahti, I., & Liebkind, K. (2011). The impact of perceived social norms, gender, and intergroup anxiety on the relationship between intergroup contact and ethnic attitudes of adolescents. *Journal of Applied Social Psychology, 41,* 1877–1899. doi:10.1111/j.1559-1816.2011.00793.x

Malti, T., Killen, M., & Gasser, L. (2012). Social judgments and emotion attributions about exclusion in Switzerland. *Child Development, 83*, 697–711. doi:10.1111/j.1467-8624.2011.01705.x

McGlothlin, H., & Killen, M. (2005). Children's perceptions of intergroup and intragroup similarity and the role of social experience. *Journal of Applied Developmental Psychology, 26,* 680–698. doi:10.1016/j.appdev.2005.08.008

McGlothlin, H., & Killen, M. (2010). How social experience is related to children's intergroup attitudes. *European Journal of Social Psychology, 40,* 625–634. doi:10.1002/ejsp.733

Mulvey, K. L., Hitti, A., Rutland, A., Abrams, D., & Killen, M. (2014). When do children dislike ingroup members? Resource allocation from individual and group perspectives. *Journal of Social Issues, 70,* 29–46. doi:10.1111/josi.12045

Naber, N. (2000). Ambiguous insiders: An investigation of Arab American invisibility. *Ethnic & Racial Studies, 23*, 37–61. doi:10.1080/014198700329123

Nesdale, D. (2004). Social identity processes and children's ethnic prejudice. In M. Bennett & F. Sani (Eds.), *The development of the social self* (pp. 219–245). New York, NY: Psychology Press.

Nesdale, D., Durkin, K., Maass, A., & Griffiths, J. (2005). Threat, group identification, and children's ethnic prejudice. *Social Development, 14,* 189–205. doi:10.1111/j.1467-9507.2005.00298.x

Nesdale, D., & Lawson, M. J. (2011). Social groups and children's intergroup attitudes: Can school norms moderate the effects of social group norms? *Child Development, 82,* 1594–1606. doi:10.1111/j.1467-8624.2011.01637.x

Nesdale, D., Maass, A., Durkin, K., & Griffiths, J. (2005). Group norms, threat, and children's racial prejudice. *Child Development, 76,* 652–663. doi:10.1111/j.1467-8624.2005.00869.x

Phinney, J. S. (2008). Bridging identities and disciplines: Advances and challenges in understanding multiple identities. *New Directions for Child & Adolescent Development, 2008*, 97–109.

Rubin, K., Bukowski, W., & Parker, J. (2006). Peer interactions, relationships, and groups. In N. Eisenberg, W. Damon, & R. M. Lerner (Eds.), *Handbook of child psychology: Social, emotional, and personality development* (Vol. 3, 6th ed., pp. 571–645). Hoboken, NJ: Wiley.

Rutland, A., Killen, M., & Abrams, D. (2010). A new social-cognitive developmental perspective on prejudice: The interplay between morality and group identity. *Perspectives on Psychological Science, 5,* 279–291. doi:10.1177/1745691610369468

Shaheen, J. G. (2003). Reel bad Arabs: How Hollywood vilifies a people. *Annals of the American Academy of Political & Social Science, 588,* 171–193. doi:10.1177/0002716203588001011

Smetana, J. G. (2006). Social-cognitive domain theory: Consistencies and variations in children's moral and social judgments. In M. Killen & J. S. Smetana (Eds.), *Handbook of moral development* (pp. 119–154). Mahwah, NJ: Erlbaum.

Stark, T. H., & Flache, A. (2012). The double edge of common interest: Ethnic segregation as an unintended byproduct of opinion homophily. *Sociology of Education, 85*, 179–199. doi:10.1177/0038040711427314

Stone, S., & Han, M. (2005). Perceived school environments, perceived discrimination, and school performance among children of Mexican immigrants. *Children and Youth Services Review, 27*, 51–66. doi:10.1016/j.childyouth.2004.08.011

Tajfel, H. (1979). Individuals and groups in social psychology. *British Journal of Social & Clinical Psychology, 18*, 183–190.

Tropp, L. R., O'Brien, T. C., & Migacheva, K. (2014). How peer norms of inclusion and exclusion predict children's interest in cross-ethnic friendships. *Journal of Social Issues, 70*, 151–166. doi:10.1111/josi.12052

Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge, UK: Cambridge University Press.

Verkuyten, M. (2004). Ethnic identity and social context. In M. Bennett & F. Sani (Eds.), *The development of the social self* (pp. 189–216). New York, NY: Psychology Press.

Verkuyten, M. (2008). Multiculturalism and group evaluations among minority and majority groups. In S. R. Levy & M. Killen (Eds.), *Intergroup attitudes and relations in childhood through adulthood* (pp. 157–172). Oxford, UK: Oxford University Press.

Wainryb, C., Shaw, L., Laupa, M., & Smith, K. R. (2001). Children's, adolescents', and young adults' thinking about different types of disagreements. *Developmental Psychology, 37,* 373–386. doi:10.1037/0012-1649.37.3.373

Child Development, September/October 2015, Volume 86, Number 5, Pages 1538–1556

The Unique and Shared Genetic and Environmental Contributions to Fear, Anger, and Sadness in Childhood

Sierra Clifford and Kathryn Lemery-Chalfant
Arizona State University

H. Hill Goldsmith
University of Wisconsin–Madison

This study examined the extent to which subordinate dimensions of negative emotionality were genetically and environmentally distinct in a sample of 1,316 twins (51% female, 85.8% Caucasian, primarily middle class, $M_{\text{age}} = 7.87$ years, $SD = .93$), recruited from Wisconsin hospital birth records between 1989 and 2004. Cholesky, independent pathway, and common pathway models were fitted for mother report, father report, and in-home observation of temperament. Although findings support the use of negative emotionality, there were heritable aspects of anger and fear not explained by a common genetic factor, and shared environmental influences common to anger and sadness but not fear. Observed fear was independent from observed anger and sadness. Distinctions support specificity in measurement when considering implications for child development.

Temperament, defined as individual differences in reactivity and regulation across affective and behavioral domains, is an important contributor to personality and adjustment in childhood and across the life span (Rothbart & Bates, 2006), with negative emotionality in particular widely linked to mood and behavior problems as well as risk for psychopathology (Lemery-Chalfant, Clifford, & Swann, in press; Nigg, 2006). Although distinctions between discrete negative emotions are well supported (Izard, 2007) and temperament research increasingly addresses proneness to discrete negative emotions as well as broad negative emotionality (e.g., Eisenberg et al., 2009), less is known about underlying relations among discrete negative emotions in childhood, including the extent to which facets of negative emotionality are genetically or environmentally related or distinct.

Behavior genetic analyses can clarify the etiological underpinnings of trait negative emotions. By considering correlations between individuals with different degrees of genetic relatedness, quantitative genetic models allow an estimation of variance within a population that can be attributed to genetic or environmental influences, including the extent to which multiple traits are related for genetic or environmental reasons (Plomin, DeFries, Knopik, & Neiderhiser, 2013). If proneness to anger, sadness, and fear represent different manifestations of an underlying proclivity toward negative emotion, these lower order traits should show considerable genetic overlap; in contrast, substantial genetic influences unique to each emotion suggest that relying on broad negative emotionality risks overlooking important behavioral distinctions. Thus, we aim to use quantitative genetics to specify the etiological overlap between fear, anger, and sadness at the subordinate level, and to test the extent to which genetic and environmental variance in each emotion can be accounted for by a single, common negative emotionality factor.

Temperament in Middle Childhood

Despite debate on the nature and underpinnings of temperament, theorists agree that temperamental differences often emerge early in development, are relatively stable over time and context, and are at least partly biologically based, with some consensus on major components of temperament (Rothbart &

This research was supported by National Institute of Mental Health Grants R01-MH59785 and R01-MH101504 and the Waisman Center Grant P30-HD003352 to Goldsmith. Special thanks to the staff and students at the Wisconsin Twin Project, especially to Nicole Schmidt for coordinating the project all these years, and to Cory Schmidt for his master coding of the observed temperament, and the participating families who generously shared their experiences.

Correspondence concerning this article should be addressed to Sierra Clifford, Department of Psychology, Arizona State University, P.O. Box 871104, Tempe, AZ 85287-1104. Electronic mail may be sent to sierra.clifford@asu.edu.

© 2015 The Authors
Child Development © 2015 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2015/8605-0016
DOI: 10.1111/cdev.12394

Bates, 2006). Some theorists emphasize neurophysiological systems underlying behavioral differences, and focus on inhibition and activation (Kagan & Fox, 2006); others (e.g., Rothbart & Bates, 2006) focus on broad domains of reactivity and regulation, organized by valence and intensity. Research supports physiological differences in approach and withdrawal, with left frontal electroencephalogram asymmetry in infancy related to approach, positive affect, and anger, and right frontal asymmetry linked to fear, withdrawal, and later social reticence and anxiety (Carver & Harmon-Jones, 2009; Fox, Henderson, Marshall, Nichols, & Ghera, 2005). However, evidence also exists for individual differences in broad negative reactivity, and factor analytic studies across multiple parent-report measures in middle childhood reveal components related to negative emotionality, self-regulation, and extraversion or approach (Rothbart & Bates, 2006). Furthermore, temperament is hierarchically structured, with higher order components made up of discrete subordinate dimensions. For instance, the negative emotionality factor of the Children's Behavior Questionnaire (CBQ) is composed of subordinate dimensions tapping anger, fear, sadness, discomfort, and low soothability (Rothbart, Ahadi, Hershey, & Fisher, 2001).

Discrete emotions have distinct antecedents, adaptive functions, and behavioral consequences (Izard, 2007). Generally, sadness may promote withdrawal and conservation of resources after loss, fear is a response to perceived threat that may lead to avoidance or to aggression if avoidance is not possible, and anger facilitates aggression and resource acquisition (Tooby & Cosmides, 2008). Unlike fear and sadness, anger motivates approach, which may explain moderate positive relations between trait anger and positive emotionality or extraversion in some studies (Carver & Harmon-Jones, 2009). Some studies also suggest distinct cortical and subcortical regions and patterns of activation underlying discrete emotions (Vytal & Hamann, 2010), but findings are mixed (Murphy, Nimmo-Smith, & Lawrence, 2003). Whether emotions represent discrete patterns of physiological and cognitive activation (Izard, 2007) or a more general combination of affect, arousal, and attribution (Russell, 1980) is unresolved, but different motivational and adaptive functions highlight the need to differentiate among emotions when predicting outcomes such as internalizing versus externalizing or aggression versus withdrawal.

Both broad negative emotionality and proneness to discrete emotions are stable over time, and broad negative emotionality predicts both internalizing and externalizing problems, although relations with externalizing problems are stronger and more consistent (Rothbart & Bates, 2006; Rothbart, Evans, & Ahadi, 2000). However, when considered separately, anger, sadness, and fear show different relations to externalizing and internalizing (Eisenberg et al., 2009). Specifically, sadness, fear, and withdrawal are stronger predictors of pure internalizing (Lemery, Essex, & Smider, 2002; Oldehinkel, Hartman, Ferdinand, Verhulst, & Ormel, 2007). Fear and behavioral inhibition in particular are consistently related to risk for anxiety and depression (Degnan & Fox, 2007; Karevold, Coplan, Stoolmiller, & Mathiesen, 2012), but may protect against externalizing (Schwartz, Snidman, & Kagan, 1996). In contrast, anger is most strongly related to externalizing (Lemery et al., 2002; Oldehinkel et al., 2007). Along with different predictive utility, there is some evidence that subordinate dimensions of negative emotionality have different genetic and environmental architecture, but this question has yet to be examined in a multivariate framework.

Quantitative Genetics Approach

Quantitative genetic research uses structural-equation-modeling-based path analysis and model fitting to decompose the total phenotypic variance in a trait into latent additive genetic (A), dominant genetic (D), shared environmental (C), and nonshared environmental (E) factors, without attempting to isolate measured genes or environmental factors contributing to variation (Plomin et al., 2013). A reflects the average effect of individual genes across the genotype, and D accounts for interaction between alleles at the same (dominance) or different (epistasis) loci. Together, A and D encompass influences that increase phenotypic similarity between individuals who are more closely genetically related. For instance, because monozygotic (MZ) twins are genetically identical, whereas dizygotic (DZ) twins share 50% of their segregating genes on average, if 100% of the resemblance in a trait was explained by additive genetic factors then phenotypic correlations between MZ twins would be expected to be approximately twice as high as DZ twin correlations. MZ correlations higher than twice DZ correlations suggest the action of D, or in some cases, rater bias. In contrast, C encompasses all influences that increase the phenotypic similarity of individuals independent of genetic relatedness. DZ correlations higher than half of MZ correlations suggest shared environmental influences, and if 100% of the resemblance in a trait were due to the shared environment, MZ and DZ correlations would be expected to be approximately equal. Finally, nongenetic factors that make individuals different from one another, including measurement error, are attributed to E. For instance, all differences between MZ twins reared in the same household can be attributed to E. Although conventions vary somewhat across studies, A, C, or E components of 30% or less are commonly described as modest, 40%–60% as moderate, and 70% as high (Plomin et al., 2013; Saudino, 2005).

Heritability of Temperament

Quantitative and molecular genetic studies support both additive genetic contributions to temperament and gene–environment interplay (Gagne, Vendlinski, & Goldsmith, 2009; Lemery-Chalfant, Kao, Swann, & Goldsmith, 2013). Heritability of temperament is moderate across childhood, generally 20%–60%, and continuity in temperament is largely genetic (Goldsmith, Buss, & Lemery, 1997; Saudino, 2005). However, parent report and observed temperament show low phenotypic convergence (Rothbart & Bates, 2006) and limited genetic or environmental overlap (Gagne & Saudino, 2010). Furthermore, some parent-report measures of temperament may overestimate heritability due to parents' tendency to contrast one DZ twin against the other (Saudino, 2003). Only studies that model contrast effects or use measures found to be less vulnerable to such biases (e.g., the Infant Behavior Questionnaire; Goldsmith et al., 1997) are reviewed here.

Parent report of negative emotionality is moderately to highly heritable in infancy, toddlerhood, and middle childhood (generally 40%–70%), with the remaining variance typically explained by nonshared environment (Goldsmith, Lemery, Buss, & Campos, 1999; Goldsmith et al., 1997; Singh & Waldman, 2010; Tackett, Waldman, Van Hulle, & Lahey, 2011). Observational measures of broad negative emotion are rare in behavioral genetic studies, but one study finds a latent factor of observed anger and experimenter rated negativity to be heritable at 14 (65%) and 20 (40%) months, with shared environment at 20 (38%) and 24 (51%; Rhee et al., 2012) months. Heritability is similar for reported fear (56%–83%), anger (66%), and sadness (71%–75%; Emde, Robinson, Corley, Nikkari, & Zahn-Waxler, 2001; Goldsmith & Lemery, 2000; Mullineaux, Deater-Deckard, Petrill, Thompson, & DeThorne, 2009), and infants' anger and fear show substantial, largely independent genetic variance (Goldsmith et al., 1999). Observed infant and toddler anger (Emde et al., 2001; Gagne & Goldsmith, 2011), and experimenter-rated anger between 4 and 8 (Deater-Deckard, Petrill, & Thompson, 2007) are moderately heritable (25%–38%). Some, but not all (e.g., Deater-Deckard et al., 2010) studies also show moderate shared environmental variance in reported and observed anger (Emde et al., 2001; Gagne & Goldsmith, 2011; Goldsmith et al., 1997) and reported social fear in early childhood (Goldsmith et al., 1997). Shared environmental effects for anger are interesting in light of links between anger and positive emotion as approach-related aspects of temperament (Carver & Harmon-Jones, 2009), as some studies show substantial shared environmental variance in positive emotionality (Goldsmith et al., 1997).

Shyness, behavioral inhibition, and stranger fear are often assessed with laboratory observations of children's interactions with unfamiliar adults or peers, and relatively consistently show high to moderate heritability and no significant role for the shared environment. Stranger fear in infancy is moderately to highly heritable according to observation (68%) and parent report of distress to novelty (58%; Goldsmith et al., 1999). Behavioral inhibition in infancy is moderately to highly heritable in interaction with unfamiliar peers (70%) and adult strangers (42%–56%) with heritability of narrow facets of inhibition ranging from 30% to 70% (DiLalla, Kagan, & Reznick, 1994; Emde et al., 1992; Robinson, Kagan, Reznick, & Corley, 1992). Finally, a study of behavioral inhibition in middle childhood finds a broad-sense heritability of 59% for exploratory behavior across nonsocial, adult and peer interaction contexts, with context-specific heritability ranging from 51% (adult) to 71% (peer; McGuire, Clifford, Fink, Basho, & McDonnell, 2003).

Genetic influences on negative emotionality are well established, but there is a need to extend quantitative genetic research on discrete emotion beyond the univariate framework, especially as subordinate facets of the same construct may have distinct patterns of genetic and environmental variance. In addition, both the stability of temperament and the range of environmental challenges (e.g., in school and peer settings; Rothbart & Bates, 2006) increase in middle childhood, but despite widespread use of negative emotionality in phenotypic research, no quantitative genetic study has examined fear, anger, and sadness together in multivariate models or considered structured observation of negative emotion past infancy or early childhood.

The Current Study

The overarching aim was to examine the genetic and environmental influences on three subordinate-level aspects of negative emotionality (anger, sadness, and fear) in middle childhood. We fit models separately for mother report, father report, and in-home observation to maximize comparisons across studies that only utilize one method. It is common to composite mother and father reports when possible, but reporter discrepancies can reflect true differences in children's behavior (De Los Reyes, Henry, Tolan, & Wakschlag, 2009). Further, mother and father reports of behavior problems show unique genetic variance, with discrepancies explained by differences in perspective as well as error (Bartels et al., 2003), and mother and father reports of temperament can differ greatly in heritability (Mullineaux et al., 2009). Given the prevalence of studies only using mother report, and the lack of research examining genetic and environmental influences on father report of temperament, examining mother and father reports separately both increases comparability to other studies and provides information on reporter differences in heritability.

The first aim was to estimate the heritability of each emotion. Regarding reporter differences, moderate to high heritability was expected for mother report of anger, sadness, and fear, whereas in-home observation was expected to be largely explained by nonshared environment, with modest to moderate genetic influences. Given limited research examining father report of temperament, no specific predictions were made. Regarding differences in heritability across emotion, fear was expected to be the most heritable emotion regardless of reporter, and modest but significant shared environment was expected for anger on the basis of past findings of shared environmental effects for anger and other approach-related traits (Emde et al., 2001; Gagne & Goldsmith, 2011; Goldsmith et al., 1997).

The second goal was to examine the genetic and environmental covariance between anger, sadness, and fear, using three increasingly restrictive models (the Cholesky decomposition, independent pathway, and common pathway models) to assess the extent to which these emotions can be explained by a common set of latent genetic and environmental factors. If the common factor provides a good fit for the data and accounts for the majority of genetic and shared environmental variance in all emotions, with residual variance largely explained by E, it supports the conceptualization of negative emotionality as a broad genetically and environmentally influenced dimension of temperament. In contrast, a poorly fitting common factor model, or substantial genetic or shared environmental factors unique to one emotion, suggest meaningful variance that is not accounted for by a negative emotionality factor. Finally, if the best fitting model is the Cholesky decomposition, it indicates that the genetic and environmental covariance among negative emotions is more complex. For example, if two distinct additive genetic factors influence sadness, one of which is shared with anger and the other with fear, the Cholesky model would be expected to provide the best fit to the data. We expected that a common, moderately heritable factor would represent the data, but also that each emotion would show important genetic and, in some cases, shared environmental variance that could not be explained by the common factor.

Method

Participants

Participants were twin children and their mothers and fathers drawn from the Wisconsin Twin Project, an ongoing longitudinal study following twins (recruited from hospital birth records in the state of Wisconsin between the years of 1989 and 2004) from toddlerhood through adolescence (Lemery-Chalfant, 2006). When twins were approximately 8 years old (M = 7.87 years, SD = .93), 933 families (1,866 twins) were selected for a follow-up assessment. Of the twins with completed zygosity assessments, 658 twin pairs (1,316 twins; 33.6% MZ, 31.0% same-gender DZ, 35.4% opposite-gender DZ; 49% male) had mother report and 1,106 had father report of temperament, with the sample size for observed emotion composites ranging from 1,166 (fear) to 1,246 (anger and sadness). The sample is largely Caucasian (85.8%). The remaining participants are African American (6.2%), Native American (1.5%), Asian (.2%), biracial (1.8%), and other (3.5%), which is representative of the population of Wisconsin. Families are largely educated and middle class, with annual family income ranging from below $10,000 to over $200,000 (mean family income between $50,001 and $60,000) and formal education ranging from 6 years to a graduate degree (M = 14.88 years for mothers and 14.43 years for fathers), with the majority of participants reporting either some college, trade or technical school, or a college degree.

Procedure

Families completed two phone interviews, mailed questionnaire packets, and a 4-hr home visit (Lemery-Chalfant, 2006). As part of the first phone interview, mothers and fathers reported on twins' temperament, along with demographic information and measures of children's behavior and environment. At the home visit, twins took part in 16 videotaped and coded episodes of the middle childhood version of the Laboratory Temperament Assessment Battery (Lab-TAB; Goldsmith, Reilly, Lemery, Longley, & Prescott, 2001). Twins were tested separately in different rooms of the house, with the same child tester administering the same episode in the same room for each twin. Time was provided between episodes for twins to rest and return to a neutral baseline state (Goldsmith et al., 2001). When tapes were coded, cotwins were coded by different raters to reduce rater bias.

Measures

Zygosity

Zygosity was assessed using mother report on the Zygosity Questionnaire for Young Twins (Goldsmith, 1991). This detailed 32-item parent-report instrument assesses physical similarities (e.g., differences in hair color, texture, and shade) and demonstrates over 95% agreement with zygosity assessed by genotyping (Price et al., 2000). Physical similarity was also observed directly, and pairs with ambiguous zygosity were genotyped (Lemery-Chalfant, 2006).

Parent-Report Temperament

Fear, anger, and sadness were assessed using mother and father reports on the CBQ (Rothbart et al., 2001). In prior studies, the CBQ has shown high longitudinal stability, and good construct and convergent validity (Rothbart et al., 2000; Rothbart et al., 2001). Ten-item scales assessing fear, anger, and sadness included questions like, "is afraid of loud noises," "gets angry when told s/he has to go to bed," and "cries sadly when a favorite toy gets lost or broken," respectively. Items are answered on a 1–7 Likert scale, with 1 being *extremely untrue of your child* and 7 being *extremely true of your child* in the past 6 months. Internal consistency at the scale level was good to adequate for both mother and father reports of anger (αs = .83, .82), sadness (αs = .69, .67), and fear (αs = .75, .75), respectively.

Observed Temperament

Observed temperament was assessed using a home-based version of the Lab-TAB designed for use with 6- to 8-year-old children (Goldsmith et al., 2001). The Lab-TAB is a standardized assessment of temperament including episodes intended to tap discrete emotional reactions under naturalistic conditions. Episodes are videotaped, range from 3 to 10 min, and are divided into coding epochs that range from 5 to 30 s. The intensity of facial, vocal, and postural displays of emotion is coded in each epoch, along with other relevant behaviors (e.g., duration of gaze aversion), as specified in the Lab-TAB manual (Goldsmith et al., 2001).

A similar preschool version of the Lab-TAB has been validated in laboratory and home settings (Gagne, Van Hulle, Aksan, Essex, & Goldsmith, 2011), showing moderate convergence between Lab-TAB composites and observer postvisit ratings (correlations ranging from .21 to .76), although correlations with parent report on the CBQ are more modest. At least 10% of episodes were double-coded by a master coder to monitor interrater agreement, and Kappas for coded variables ranged from .79 to .89 for fear, .70 to .93 for sadness, and .63 to .84 for anger. Although kappa-indexed agreement was good, we also standardized scores within coder, then rescaled them back onto the raw scores of a referent coder (usually the coder who coded the greatest number of episodes) to further reduce any coder differences (see first author for details).

Selected episodes: Anger and sadness. Observed anger and sadness were assessed using four Lab-TAB episodes that regularly evoke both sadness and anger responses from children. In "I'm Not Sharing," the child tester unequally shares a bag of candy with the child. In "Impossibly Perfect Stars," the child is asked to repeatedly draw stars by a child tester, who responds by critiquing each star and asking the child to try again. In "Transparent Box," the child is presented with several appealing toys in a locked transparent box, along with a set of keys that cannot open the box, and in "Wrong Gift," instead of a preferred gift, the child is given a box containing an unappealing gift that the child has previously ranked as his or her least favorite. Stars, Not Sharing, and Transparent Box are coded in 10-s epochs, and Wrong Gift in 5-s epochs, with two 4-min episodes (Stars and Transparent Box) and two variable-length episodes (Not Sharing and Wrong Gift). Latency in seconds, and intensities of facial, bodily, and vocal expressions of anger and sadness on 0–2 or 0–3 scales, were coded in accordance with the Lab-TAB manual (Goldsmith et al., 2001), and lower order parameters included in each episode-level composite are summarized in Table 1. In each episode, mean intensity scales were coded for facial and bodily anger and sadness. Latency to facial or bodily anger and sadness was coded in all episodes except Wrong Gift, which included a discrete emotion-eliciting event. Other parameters coded in each episode differed. Specifically, vocal expression was coded in Stars and Not Sharing. In addition, during the coding for the current study, we combined facial and bodily scales for anger and sadness in Not Sharing into scales including both facial and bodily emotion due to low frequency of these responses. Thus, Stars included speed and vocal, bodily and facial intensity; Not Sharing included speed to first bodily or facial response, speed to first vocalization, bodily and facial intensity, and vocal intensity; Transparent Box included speed, bodily intensity and facial intensity; and Wrong Gift included bodily and facial intensity.

Selected episodes: Fear. Fear was assessed with two episodes, "Scary Mask" and "Storytelling." In Scary Mask, coded in 15-s epochs, the child interacts with a friendly stranger wearing a frightening mask, beginning when the child first notices the stranger's face and continuing as the stranger takes off the mask, begins a conversation with the child, and finally asks the child to touch and wear the mask. In Storytelling, coded in 10-s epochs, the child is asked to give a speech about the previous day's events in front of an audience of multiple child testers. The episode includes at least one prompt by the child tester (e.g., "Is there anything else you would like to tell us?"). In both episodes, variables are coded on scales of 0–3 in accordance

Table 1
Summary of Episode-Level Composites

Emotion	Episode	Parameters included in final composite	Scale	Skew	Kurtosis	α	*r*
Anger	Stars	Speed (facial/bodily)	Time (s)	−1.11	−0.07	.68	
		Mean facial intensity	0–2	1.30	1.93		
		Mean bodily intensity	0–2	2.02	5.25		
		Mean vocal intensity	0–2	2.46	6.87		
	Not Sharing	Speed (facial/bodily)	Time (s)	−1.16	−0.26	.76	
		Speed (vocal)	Time (s)	0.46	−1.63		
		Mean facial/bodily intensity	0–3	1.50	2.80		
		Mean vocal intensity	0–3	2.05	0.07		
	Transparent Box	Speed (facial/bodily)	Time (s)	−0.12	−1.45	.79	
		Mean facial intensity	0–2	2.43	7.32		
		Mean bodily intensity	0–2	2.73	7.43		
	Wrong Gift	Mean facial intensity	0–2	2.01	5.39		.60
		Mean bodily intensity	0–2	2.60	9.81		
Sadness	Stars	Speed (facial/bodily)	Time (s)	−0.18	−1.11	.71	
		Mean facial intensity	0–2	3.01	11.93		
		Mean bodily intensity	0–2	2.14	5.78		
		Mean vocal intensity	0–2	1.58	2.03		
	Not Sharing	Speed (facial/bodily)	Time (s)	−1.31	0.72		.49
		Mean facial/bodily intensity	0–3	1.86	4.48		
	Wrong Gift	Mean facial intensity	0–2	1.72	2.72		.61
		Mean bodily intensity	0–2	1.72	2.93		
Fear	Scary Mask	Speed (facial/bodily)	Time (s)	−2.94	7.07	.63	
		Mean facial intensity	0–3	0.91	0.11		
		Mean bodily intensity	0–3	0.23	−0.58		
	Storytelling	Speed (facial/bodily)	Time (s)	−0.72	−1.30		.42
		Mean bodily intensity	0–3	1.02	−0.18		

Note. Intensity of approach, avoidance, and vocal expression in Scary Mask; intensity of avoidance in Storytelling; and speed and intensity of vocal sadness in Not Sharing were coded but not included in episode-level composites due to low correlations with other parameters. Facial intensity was coded in Storytelling but not included due to low frequency of responses above 0. No sadness composite was formed for Transparent Box due to low frequency of responses above 0. Square root transformations were applied to parameters with skewness > 2.00 or kurtosis > 7.00. Parameters were standardized using *z*-score transformations, and unit-weighted mean composites were formed.

with the Lab-TAB manual (Goldsmith et al., 2001) and include latency to first fear in seconds, facial and bodily intensity, and intensity of avoidance (summarized in Table 1). Intensity of approach and vocal intensity were also coded in Scary Mask but not Storytelling.

Lab-TAB composite formation

Lower order composites. Latency scores were winsorized to 3 *SD*, then reverse coded to transform them to speed scores. Mean intensity scores were computed across all epochs for facial, bodily, and vocal emotion in each episode. Three parameters (mean intensity of facial fear in Storytelling, and bodily and facial sadness in Transparent Box) were discarded due to low frequency of scores above zero (< 30%). Square root transformations were used for lower order parameters exceeding recommended cutoffs for skewness and kurtosis (2.00 and 7.00, respectively; Muthén & Kaplan, 1985) prior to observed composite formation (see Table 1 for skewness and kurtosis). *Z*-score transformations were used to standardize lower order speed and intensity parameters prior to higher order composite formation. Mean composites for anger, sadness, and fear were formed from speed and mean intensity scores on the basis of zero-order correlations, with variables only included in an episode-level composite if they were moderately and significantly correlated with at least one other lower order parameter ($r > .30$). For anger, episode-level composites included all available emotion parameters for Stars (speed, vocal intensity, facial intensity, bodily intensity), Not Sharing (bodily and facial speed, vocal speed, bodily and facial intensity, vocal intensity), Transparent Box (speed, bodily intensity, facial intensity), and Wrong Gift (bodily and expressed intensity). Correlations across all lower order anger parameters within episode ranged from low to moderate (.217–.677), but each parameter was correlated above $r = .30$ with at least one other parameter. For sadness, all available emotion parameters were used for Stars and Wrong Gift (*r*s from .203 to .698), but for Not Sharing, vocal parameters were not highly correlated with speed and intensity of facial and bodily sadness, and the final sadness composite for Not Sharing was formed from speed and intensity of bodily and facial sadness ($r = .492$). The final fear composite for Scary Mask included speed, facial, and bodily fear (*r*s from .340 to .366), but vocal intensity and intensity of avoidance were excluded due to low correlations with other parameters. The final composite for Storytelling included only speed and bodily intensity ($r = .420$), with facial fear excluded due to low variability and intensity of avoidance excluded due to low correlations.

Higher order cross-episode composites. Episode-level anger composites for Stars, Not Sharing, and Transparent Box, and sadness composites for Stars and Not Sharing were standardized using *z*-score transformations and used to form cross-episode mean composites of anger and sadness, respectively. Correlations among episode-level anger composites were moderate, with *r*s ranging from .249 (Not Sharing and Transparent Box) to .311 (Stars and Not Sharing), whereas for sadness, only the cross-episode composites for Stars and Not Sharing were aggregated ($r = .174$). Fear assessed in Storytelling and Scary Mask were moderately correlated and composited ($r = .262$).

Data Analysis Plan

Analyses for Aim 1: Univariate ACE Models

The latent A, C, and E factors in the classic univariate biometrical ACE model are estimated from observed variances and covariances of MZ and DZ twins' scores on a single trait. The ACE model is a multigroup structural equation model, with covariances modeled differently for MZ (who share 100% of their segregating DNA) and DZ (who share on average 50%) twins. C is fully shared between cotwins; thus, the correlation between latent C factors is fixed to 1.0. Due to model identification limitations with twin data, it is not possible to estimate both C and D factors within the same model, but it is possible to fit an alternate ADE model, which accounts for nonadditive genetic effects (MZ twins share 100% of nonadditive genetic effects, whereas DZ twins inherit the same alleles at a locus 25% of the time). The E factor is uncorrelated for both MZ and DZ twins.

We fit univariate ACE or ADE models estimating genetic and environmental influences on sadness, anger, and fear separately for mother report, father report, and in-home observation using the statistical program OpenMX (Boker et al., 2011), an R-based program that uses maximum likelihood estimation. After the full models were fit, nested models were tested by systematically dropping parameters, and the fit of reduced models was compared with that of the full model using the −2 log likelihood chi-square test of fit to find the most parsimonious solution. A nonsignificant difference in fit implies that the reduced model accounts for the observed data as well as the full model, whereas a significant loss of fit indicates the dropped parameter is required to adequately reproduce the data. Because E contains measurement error, it is always retained, and D is not estimated without A, as it is unlikely that all genetic influences are interactive, with no additive effects. Because variance components are interdependent and bounded at zero, standard error based confidence limits may be inaccurate. We obtained 95% confidence limits for variance components using the likelihood-based confidence interval option in OpenMx, which computes nonsymmetric confidence limits by changing the value of the selected parameter until the model fit is reduced by a specified amount, while the other estimated parameters are allowed to vary freely (OpenMx Development Team, 2011). We used the default reduction in fit, a change in −2 log likelihood of 3.841459, corresponding to 95% confidence on a chi-square distribution with 1 *df* (OpenMx Development Team, 2011).

Analyses for Aim 2: Multivariate ACE Models

Cholesky decompositions. Just as univariate models decompose the variance in a single trait into genetic and environmental factors, Cholesky decompositions (Figure 1a) parse the covariance between traits into latent A, C or D, and E factors, allowing an estimation of the extent to which genetic and environmental influences on one trait are shared with other traits, regardless of trait heritability. In a trivariate Cholesky decomposition, the first set of latent factors (A1, C1, and E1) encompasses genetic, shared and nonshared environmental influences on the first trait, which may be shared with the second and third traits. The second set of latent factors (A2, C2, and E2) represents influences on the second trait that are independent of the first, but may be shared with the third. The third set of latent factors represents influences unique to the third trait. Thus, the order of traits is

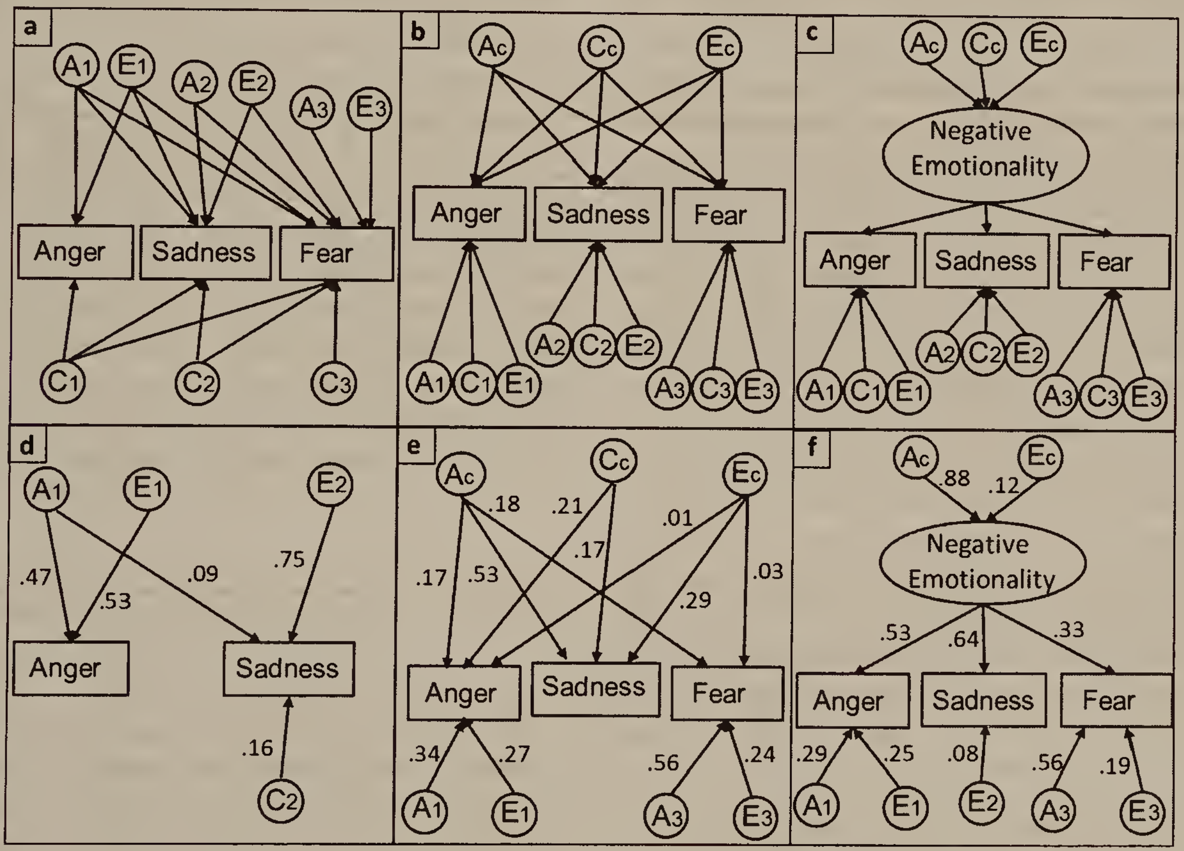

a
A1
E1
A2
E2
A3
E3
Anger
Sadness
Fear
C1
C2
C3
b
Ac
Cc
Ec
Anger
Sadness
Fear
A1 C1 E1
A2 C2 E2
A3 C3 E3
c
Ac
Cc
Ec
Negative Emotionality
Anger
Sadness
Fear
A2 C2 E2
A1 C1 E1
A3 C3 E3
d
A1
E1
E2
.47
.53
.09
.75
Anger
Sadness
.16
C2
e
Ac
Cc
Ec
.18
.21
.01
.17
.17
.53
.03
.29
Anger
Sadness
Fear
.34
.27
.56
.24
A1
E1
A3
E3
f
Ac
Ec
.88
.12
Negative Emotionality
.53
.64
.33
Anger
Sadness
Fear
.29
.25
.08
.56
.19
A1
E1
E2
A3
E3

Figure 1. Top: Example trivariate Cholesky decomposition (a), independent pathway model (b), and common pathway model (c), showing one twin only for simplicity. Latent A_1, C_1, and E_1 factors in the Cholesky decomposition represent additive genetic, shared environmental, and nonshared environmental influences on anger, which may also be shared with sadness and fear. A_2, C_2, and E_2 factors represent A, C, and E influences on sadness that are independent of anger but may be shared with fear, and A_3, C_3, and E_3 factors represent influences unique to fear. In the independent pathway model, covariance among emotions is fully accounted for by a single set of common additive genetic (A_C) shared environmental (C_C), and nonshared environmental (E_C) factors, but the relative loadings on each phenotype may vary. Unique additive genetic, shared environmental, and nonshared environmental influences are also estimated (e.g., A_1, C_1, and E_1 are unique to anger). In the common pathway model, covariance is fully accounted for by a single common factor (i.e., negative emotionality), decomposed into additive genetic (A_C), shared environmental (C_C), and nonshared environmental (E_C) factors. Bottom: Final best fitting models for observed (d), mother report (e), and father report (f). Standardized parameter estimates are reported.

relevant to the interpretation of each set of factors, but does not affect model fit or the significance of shared A, C, or E variance between any two traits. We selected anger as the first trait and fear as the third because anger and fear were not expected to be strongly related, and we expected sadness to share covariance with both traits.

Independent and common pathway models. The independent pathway model (Figure 1b; Martin & Eaves, 1977) assumes that the same A, C, and E factors account for the covariance among anger, sadness, and fear, and allows the magnitude of paths from these shared factors to each emotion to differ, such that, for example, covariance between anger and sadness may be explained largely by C, and covariance between sadness and fear by A. Because the independent pathway model is not nested within the Cholesky decomposition, we used Akaike's information criterion (AIC; Akaike, 1987) to compare the fit of these two models, with lower values indicating a model with more support. The common pathway model (Figure 1c; McArdle & Goldsmith, 1990) is nested within the independent pathway model and assumes a common negative emotionality factor, which itself has single estimates of A, C, and E, with each emotion loading on this factor. Thus, unless residuals are allowed to correlate, the common pathway model requires that any A, C, and E factors that explain covariance between two emotions must also be shared with the third. Genetic and environmental variance in each trait that is independent of the common factor is also estimated. To the extent that covariance between anger, sadness, and fear is explained by a common set of A, C, and E factors, and the structure of these factors is similar for each emotion, the common pathway model provides a good fit to the data.

Results

Preliminary Analyses

Means and standard deviations are given in Table 2; all parent-report scales and observed epi-

Table 2
Means, Standard Deviations, Ranges, Zero-Order Correlations, and Twin Intraclass Correlations

	1	2	3	4	5	6	7	8	9	10	11	MZ	SSDZ	OSDZ
1. Anger (M) –	—	.56***	.26***	.62***	.36***	.15***	.78***	.49***	.05	.06*	−.03	.69	.48	.53
2. Sadness (M)		—	.42***	.33***	.48***	.24***	.82***	.45***	.00	.03	.01	.73	.54	.40
3. Fear (M)			—	.10***	.20***	.58***	.74***	.40***	−.07*	.01	−.04	.76	.32	.31
4. Anger (F)				—	.58***	.21***	.45***	.78***	.09*	.07*	−.01	.71	.32	.40
5. Sadness (F)					—	.35***	.43***	.81***	.04	.08*	−.03	.73	.47	.40
6. Fear (F)						—	.43***	.71***	−.04	.05	−.03	.81	.44	.31
7. NE (M)							—	.57***	−.01	.05	−.02	.77	.53	.53
8. NE (F)								—	−.04	.09*	−.03	.81	.46	.43
9. Anger (O)									—	.21*	.13*	.47	.31	.21
10. Sadness (O)										—	.01	.26	.34	.12
11. Fear (O)											—	.50	.45	.45
Male														
M	4.63^{a}	3.84	3.79	4.56^{a}	3.77	3.71	4.08	4.01	0.12^{a}	0.02	−0.08^{b}			
SD	0.92	0.76	0.98	0.85	0.71	0.92	0.67	0.62	0.77	0.80	.80			
Minimum	1.50	1.70	1.40	1.70	1.70	1.30	1.93	1.93	−1.65	−1.90	−1.96			
Maximum	7.00	5.90	6.60	6.71	5.89	6.22	6.16	5.67	2.33	2.36	2.02			
Female														
M	4.38	3.99	3.98	4.31^{a}	3.89	3.96	4.12	4.05	−0.11^{a}	−0.02	0.07^{b}			
SD	0.97	0.80	0.99	0.89	0.71	0.90	0.74	0.65	0.72	0.77	0.77			
Minimum	1.00	1.40	1.38	1.40	1.40	1.80	1.38	1.74	−1.83	−1.65	−1.89			
Maximum	6.90	6.27	6.70	6.80	5.97	6.59	6.22	5.94	2.33	2.36	1.88			

Note. M = mother report; F = father report, O = in-home observation; NE = negative emotionality, a mean composite of anger, sadness, and fear scales that does not include any independent assessment. Observational assessments are mean composites of standardized scales and thus have means close to zero. Twin intraclass correlations are reported on the right. MZ = monozygotic; SSDZ = same gender dizygotic; OSDZ = opposite-gender dizygotic twin pairs.

^{a}Males had significantly higher mother report, father report, and observed anger. ^{b}Females had significantly higher observed fear. All other gender differences in means were nonsignificant.

*$p < .05$. ***$p < .001$.

sode-level and cross-episode composites were normally distributed. Phenotypic correlations between mother and father reports and observed anger, sadness, and fear are given in Table 2. Because twins are clustered in families, correlations were run in MPlus using the type=complex and cluster options (Muthén & Muthén, 1998–2011). As expected, correlations between emotions were high to moderate for both parent reports, whereas correlations between observed emotions were more modest, and convergence between parent report and observed emotions was low. Age and family socioeconomic status (SES) residualized scores were used for twin modeling to account for these effects on ACE estimates, as is standard practice to reduce potential biases when it is infeasible to incorporate additional covariates (McGue & Bouchard, 1984).

Twin Intraclass Correlations

Table 2 reports twin intraclass correlations (ICCs). Mother report of anger and sadness and father report of sadness show DZ correlations higher than half the MZ correlations, suggesting both additive genetic and shared environmental factors, whereas MZ correlations approximately twice as high as DZ correlations for mother report of fear and father report of anger and fear suggest additive genetic and nonshared environmental variance. ICCs for observed emotion show evidence of heritability for anger, with shared environmental factors likely to be important for observed sadness and fear. ADE models were tested for mother and father reports of fear, as DZ correlations lower than half MZ correlations suggest nonadditive effects are plausible.

Quantitative Genetic Analyses

Saturated Models

Saturated models were fit to test for gender differences, rater contrast, and assimilation effects. Fully saturated multigroup models freely estimating means, variances and covariances for male MZ, female MZ, male DZ, female DZ, and opposite-gender DZ twins were tested against a series of models constraining means and variances to be equal across twin pairs and zygosity groups, and constraining means, variances and covariances to be equal across gender. Higher phenotypic variance in DZ relative to MZ zygosity groups indicates that parents may be contrasting the DZ twins against each other or inflating the similarity of MZ twins (assimilation effects), whereas lower DZ variance indicates possible sibling cooperation or imitation effects (Neale & Cardon, 1992; Saudino, 2003). In each case except for anger and observed fear, means, variances and covariances could be equated across gender, and means and variances across zygosity. For mother report, father report and observed anger, males had higher means than females, and for observed fear, females had higher means than males, but there were no significant gender differences in variances or covariances between same-gender male, same-gender female, and opposite-gender groups for any variable. In all models tested, variances could be constrained to be equal across MZ and DZ twins, providing no evidence of rater contrast effects.

Univariate ACE and ADE Models

Standardized estimates of A, C (or D), and E factors for each emotion, and fit statistics of full and best fitting reduced models are given in Table 3.

According to the best fitting models for both mother and father reports, all emotions were moderately to highly heritable, with A influences lowest for mother report anger (.45) and sadness (.45) and highest for father report fear (.80). Significant C was found for mother report anger (.26) and sadness (.27), but not fear, which was best explained by an AE model with high A (.74). For father report, anger (.74) and fear (.80) were highly heritable, with no significant C, but although father report sadness was moderately heritable (.53), the C factor (.19) could not be dropped without a significant loss of fit. Similar to parent report, observed anger had moderate additive genetic (.47) and nonshared environmental (.53) variance. For observed sadness and fear, the CE model fit the data best, with modest but significant C for sadness (.22) and moderate C for fear (.46). For all observed emotions, the majority of variance was explained by E.

Genetic and Environmental Covariance Across Emotion

Two trivariate Cholesky decompositions were fit for anger, sadness, and fear, one for mother report and one for father report. It was not possible to fit a trivariate Cholesky decomposition using observed emotion, because fear was not sufficiently correlated with anger or sadness ($r < .10$), but a bivariate Cholesky decomposition was fit examining the covariance between anger and sadness. Fit statistics for the full and best fitting reduced models, as well

Table 3
Univariate ACE/ADE Model Fit and Parameter Estimates

Scale	Model	−2LL	*df*	Δ−2LL	Δ*df*	*p*	AIC	A	C/D	E
Mother report										
Anger (M)	**ACE**	**3,200.83**	**1,296**	—	—	—	**608.83**	**.45 (.28–.62)**	**.26 (.11–.41)**	**.29 (.24–.35)**
	AE	3,212.3	1,297	11.47	1	< .001	618.3			
	CE	3,224.38	1,297	23.55	1	< .001	630.38			
	E	3,467.2	1,298	266.37	2	< .001	871.2			
Sadness (M)	**ACE**	**2,735.37**	**1,296**	—	—	—	**143.37**	**.45 (.28–.62)**	**.27 (.11–.42)**	**.28 (.24–.35)**
	AE	2,746.55	1,297	11.18	1	< .001	152.55			
	CE	2,759.46	1,297	24.09	1	< .001	165.46			
	E	3,013.49	1,298	278.12	2	< .001	417.49			
Fear (M)	ADE	3,338.77	1,296	—	—	—	746.77	.44 (.09–.79)	.31 (.00–.67)	.25 (.09–.30)
	AE	**3,342.11**	**1,297**	**3.34**	**1**	**0.07**	**748.11**	**.74 (.65–.84)**	—	**.26 (.22–.31)**
	E	3,561.64	1,298	222.87	2	< .001	965.64			
Father report										
Anger (F)	ADE	2,589.2	1,088	—	—	—	413.2	.71 (.34–.84)	.02 (.00–.39)	.27 (.22–.33)
	AE	**2,589.21**	**1,089**	**0.01**	**1**	**0.92**	**411.21**	**.73 (.62–.84)**	—	**.27 (.23–.33)**
	E	2,770.14	1,090	180.94	2	< .001	590.14			
Sadness (F)	**ACE**	**2,101.4**	**1,088**	—	—	—	**−74.6**	**.53 (.34–.73)**	**.19 (.01–.36)**	**.28 (.23–.36)**
	AE	2,105.87	1,089	4.47	1	0.03	−72.13			
	CE	2,128.32	1,089	26.92	1	< .001	−49.68			
	E	2,321.43	1,090	220.03	2	< .001	141.43			
Fear (F)	ADE	2,624.4	1,088	—	—	—	448.4	.77 (.40–.90)	.03 (.00–.38)	.20 (.17–.25)
	AE	**2,624.42**	**1,089**	**0.02**	**1**	**0.89**	**446.42**	**.80 (.70–.90)**	—	**.20 (.17–.25)**
	E	2,873.52	1,090	249.12	2	< .001	693.52			
Observed										
Anger (O)	ACE	2,652.21	1,231	—	—	—	190.21	.39 (.12–.57)	.06 (.00–.27)	.55 (.46–.66)
	AE	**2,652.58**	**1,232**	**0.37**	**1**	**0.54**	**188.58**	**.47 (.36–.57)**	—	**.53 (.46–.63)**
	CE	2,660.1	1,232	7.89	1	< .001	196.1			
	E	2,730.44	1,233	78.23	2	< .001	264.44			
Sadness (O)	ACE	2,831.78	1,230	—	—	—	371.78	.06 (.00–.36)	.18 (.00–.30)	.76 (.64–.87)
	AE	2,834.29	1,231	2.51	1	0.11	372.29			
	CE	**2,831.94**	**1,231**	**0.16**	**1**	**0.69**	**369.94**	—	**.22 (.14–.31)**	**.78 (.70–.87)**
	E	2,862.82	1,232	31.04	2	< .001	398.82			
Fear (O)	ACE	2,589.27	1,164	—	—	—	261.27	.00 (.00–.23)	.46 (.27–.56)	.54 (.45–.61)
	AE	2,610.58	1,165	21.32	1	< .001	280.58			
	CE	**2,589.27**	**1,165**	**0**	**1**	**.99**	**259.27**	—	**.46 (.37–.56)**	**.54 (.48–.61)**
	E	2,722.58	1,166	133.31	2	< .001	390.58			

Note. A, C, D, and E are standardized squared parameter estimates for additive genetic, common environment, dominant genetic, and nonshared environment factors, respectively. Standardized confidence intervals are reported in parentheses. The most parsimonious final model is indicated in bold. −2LL = −2 log likelihood; Δ = change; AIC = Akaike's information criterion; M = mother report; F = father report; O = observer report.

as unstandardized and standardized A, C, and E squared parameter estimates, are given in Table 4.

For observed anger and sadness, the best fitting model was an AE-ACE model (Figure 1d). The trait correlation between anger and sadness was fully explained by A, but this A factor explained 47% of the variance in anger and only 9% of the variance in sadness. Sadness had significant C (.16), but both emotions were largely explained by emotion-specific E. For mother report, the best fitting model was an ACE-ACE-AE model with both shared and emotion-specific A. One set of A factors explained 47% of the variance in anger, 17% of the variance in sadness, and 9% of the variance in fear. A second set of A factors independent of anger explained 36% of the variance in sadness and 10% of the variance in fear, but fear was largely explained by emotion-specific A (.55). Furthermore, C was significant for anger (.24) and sadness (.18) but not fear, and all C influences could be explained by a single factor fully shared across emotion. A small proportion of E was shared between anger and sadness (.01)

Table 4

Multivariate Model Fit and Parameter Estimates

Model	−2LL	*df*	Δ−2LL	Δ*df*	*p*	AIC
Best fitting full and final models (observed)						
Bivariate Cholesky (full)	5,444.50	2,460	—	—	—	524.50
Bivariate Cholesky (final)	**5,445.51**	**2,464**	**1.17**	**4**	**.88**	**517.67**
Model comparisons (mother report)						
Trivariate Cholesky (full)	8,700.50	3,882	—	—	—	936.50
Trivariate Cholesky (final)	8,709.62	3,887	9.11	5	.10	935.62
IPM (full)	8,700.64	3,882	—	0	—	936.64
IPM (final)	**8,709.93**	**3,888**	**9.29**	**6**	**.16**	**933.93**
CPM (full)	8,722.56	3,886	21.92	4	< .001	950.56
CPM (final)	8,722.56	3,890	0	4	.99	942.56
Model comparisons (father report)						
Trivariate Cholesky (full)	6,873.51	3,258	—	—	—	357.51
Trivariate Cholesky (final)	6,885.82	3,266	12.31	8	.14	353.82
IPM (full)	6,876.27	3,258	—	0	—	360.27
IPM (final)	6,889.34	3,267	13.07	9	.16	355.34
CPM (full)	6,881.93	3,262	5.67	4	.23	357.93
CPM (final)	**6,885.13**	**2,367**	8.87	9	**.45**	**351.13**

Model	Scale	A1	C1	E1	A2	C2	E2
Parameter estimates for best fitting final models							
Observed							
Bivariate Cholesky	Anger	.26/.47 (.37–.58)	—	.28/.53 (.45–.62)			
	Sadness	.06/.09 (.05–.15)	—	—	—	.09/.16 (.08–.23)	.45/.75 (.67–.85)

Model	Scale	Ac	Cc	Ec	Au	Cu	Eu
Mother report							
Independent pathway model	Anger	.15/.17 (.09–.28)	.17/.21 (.10–.33)	.01/.01 (.003–.03)	.28/.34 (.23–.42)	—	.23/.27 (.23–.33)
	Sadness	.31/.53 (.38–.68)	.10/.17 (.06–.30)	.17/.29 (.25–.35)	—	—	—
	Fear	.17/.18 (.11–.29)	—	.02/.03 (.01–.05)	.56/.56 (.45–.66)	—	.22/.24 (.20–.29)

	Scale	Ac	Cc	Ec	Au	Cu	Eu	λ
Father report								
Model common pathway model	NE	.88 (.81–.94)	—	.12 (.06–.19)				
	Anger				.29/.38 (.30–.46)	—	.18/.25 (.20–.30)	.53
	Sadness				—	—	.08/.16 (.10–.22)	.64
	Fear				.56/.68 (.59–.78)	—	.16/.19 (.16–.24)	.33

Note. See Figure 1 for clarification of A, C, and E common (e.g., A1, Ac) and unique (e.g., A3, Au) latent factors. −2 log likelihood values and *p* values are not reported for comparisons between the Cholesky decomposition and the independent pathway model because these models are not nested. A, C, and E values are first reported as unstandardized squared parameter estimates, followed by standardized parameter estimates, with confidence intervals for standardized parameters in parentheses. The most parsimonious final model is indicated in bold. −2LL = −2 log likelihood; Δ = change; *p* = probability; AIC = Akaike's information criterion; A = additive genetic; C = common environment; D = dominant (nonadditive) genetic; E = nonshared environment; IPM = independent pathways model; CPM = common pathways model; NE = latent negative emotionality; λ = factor loadings of each phenotype on NE; Ac, Cc, and Ec = A, C, and E variance common to all phenotypes; Au, Cu, and Eu = unique A, C, and E variance.

and between sadness and fear (.02), but E was largely emotion-specific. For father report, the best fitting model was an AE-AE-AE model, with both shared and emotion-specific A. Anger was highly heritable (.72), and A influences on anger also explained substantial variance in sadness (.35) and modest variance in fear (.05). A second set of A factors explained 38% of the variance in sadness and 12% of the variance in fear, but fear was largely explained by emotion-specific A (.63). As with mother report, E was largely emotion specific, with anger and sadness sharing a small E factor (.02).

Independent and Common Pathway Models

Independent pathway and common pathway models were tested for mother and father reports, but not for observed emotion. This is because observed fear was insufficiently correlated with either observed anger ($r = .134$) or sadness ($r = .014$) to support a three-indicator model, as is recommended for model identification (McArdle & Goldsmith, 1990), and fixing one path for model identification can distort estimates of genetic and environmental variance. Table 4 summarizes the fit of the full and most reduced independent pathway models for mother and father reports in comparison to the corresponding full and most reduced Cholesky decompositions, as well as parameter estimates and comparisons of fit between the full and reduced independent pathway models and the nested common pathway models. For mother report, the independent pathway model did not fit significantly worse than the Cholesky decomposition (ΔAIC = .14, $\Delta df = 0$ with the same number of parameters estimated), suggesting that the genetic and environmental covariance between these emotions can be represented by a single set of ACE factors. However, the common pathway model led to a significant loss of fit relative to the independent pathway model, $\Delta\chi^2(4) = 21.92$, $p < .001$, ΔAIC = 13.92. In the final independent pathway model (Figure 1e), it was necessary to retain shared A and E factors for all emotions, and a shared C factor for anger and sadness, but the path from the shared C factor to fear could be dropped. Interestingly, for sadness, all A (.53), C (.17), and E (.29) variance was fully shared with other emotions, and for anger, all C (.21) was explained by the shared C factor. However, anger and fear had both shared and emotion-specific A, with 17% of the variance in anger and 18% of the variance in fear explained by the shared A factor, and 34% of the variance in anger and 56% of the variance in fear explained by emotion-specific A. It was necessary to retain paths from the common E factor to both anger and fear, but the variance explained by this factor was small (.01–.03), with 27% of the variance in anger and 24% of the variance in fear explained by emotion-specific E factors.

For father report, neither the independent pathway model (ΔAIC = 2.76, $\Delta df = 0$) nor the common pathway model led to significant loss of fit relative to the full Cholesky decomposition (ΔAIC = .42, $\Delta df = 4$), and the common pathway model did not fit significantly worse than the full independent pathway model, $\Delta\chi^2(4) = 5.67$, $p = .23$, ΔAIC = −2.34, suggesting that a single negative emotionality factor provides an adequate representation of the data (see Table 4). This factor (Figure 1f) was highly heritable (.88), with the remainder of variance accounted for by E. However, substantial emotion-specific A was needed to explain both anger (.38) and fear (.68), and all emotions had significant emotion-specific E ranging from 16% for sadness to 25% for anger. Factor loadings on the common factor were higher for sadness (.64) and anger (.53) than fear (.33), which seemed to be relatively more independent of the other emotions.

Discussion

We used quantitative genetic modeling to characterize the genetic and environmental underpinnings of anger, sadness, and fear in middle childhood, with the aim of examining the extent to which these emotions represent a broad, genetically and environmentally influenced predisposition toward negative emotionality. Results support negative emotionality as a coherent, genetically influenced trait, but provide clear evidence for genetic influences on anger and fear that are not shared with other emotions, and thus also support an emotion-specific approach.

Heritability of Negative Emotions in Middle Childhood

Because genetically informed temperament research has focused on infancy and early childhood, one aim of this study was to describe the univariate heritability of anger, sadness, and fear in middle childhood. Consistent with past research (Gagne et al., 2009; Saudino, 2005), mother and father reports of all emotions were moderately to highly heritable, with the highest heritability for mother and father reports of fear, and the lowest for mother report of anger and sadness and father report of sadness. Some unanticipated findings did emerge, most notably the presence of shared environmental variance in dimensions of emotion that were largely explained by genetic factors in past studies (Goldsmith et al., 1997; Mullineaux et al., 2009). Specifically, we observed shared environmental effects at the univariate level for mother report, father report, and observed sadness, as well as negligible heritability and moderate shared environment for observed fear. However, although past studies find modest shared environment for observed and mother report anger (Emde et al., 2001; Gagne & Goldsmith, 2011; Goldsmith et al., 1997), we only found shared environmental variance in mother report anger, with father report highly and observer report moderately heritable. Anger is theoretically and empirically linked to approach-related aspects of temperament such as positive emotionality (e.g., Carver & Harmon-Jones, 2009; Deater-Deckard et al., 2010), which is also less heritable (Goldsmith et al., 1997; Goldsmith et al., 1999; Saudino, 2005). However, our findings are consistent with some literature, as not all studies of anger find shared environmental effects (Deater-Deckard et al., 2010; Mullineaux et al., 2009).

Aspects of temperament related to fear, shyness, or inhibition are highly heritable when assessed in laboratory interaction with peers or adults (e.g., DiLalla et al., 1994; Goldsmith et al., 1999), perhaps because unfamiliar situations and stranger interaction facilitate rather than suppress a child's predisposition toward shyness or withdrawal. Our use of in-home rather than laboratory observation may have contributed to differences between this and prior studies regarding observed fear, as the home is likely a more comfortable context for some shy or fearful children. Because behavior genetic research consistently finds that heritability of temperament tends to increase rather than decrease with age (Saudino, 2005), the higher shared environmental variance in the current study relative to past research in younger samples is unlikely to be due to age differences. However, past research in middle childhood has either examined lower order dimensions of temperament in a small sample, with limited power to detect shared environmental effects (Mullineaux et al., 2009), or examined broad negative emotionality over wide age ranges from childhood to adolescence (Singh & Waldman, 2010; Tackett et al., 2011), which may underestimate emotion-specific shared environmental influences such as we found for sadness.

Finally, we found high nonshared environmental effects for observed temperament, particularly sadness. Nonshared environmental effects often range from 40% to 80% for observed temperament (Saudino, 2005), and our findings are comparable (54%–76%). Although this variance may index meaningful contributions of the child's environment, it may also index measurement error, as well as the sensitivity of behavioral observation to daily variation and context. Furthermore, due to the use of separate coders for each twin in the current study, rater differences may have lead to higher estimates of nonshared environmental variance.

Genetic and Environmental Influences on Covariance Across Emotions

The second aim was to test whether genetic and environmental influences shared across emotions could be accounted for by a common factor, or whether more specificity would be required to explain relations among emotions. Despite evidence for distinct neural and biological systems involved in discrete emotions (Vytal & Hamann, 2010), many developmental theories assume that genetic and environmental influences on anger, sadness, and fear act through a biologically based, higher order negative emotionality dimension (Rothbart & Bates, 2006). We tested two nested models, the independent pathway model and the common pathway model, both of which posit that a single set of shared genetic and environmental factors is able to explain covariance among emotions. The independent pathway model allows the loadings of the three emotion measures to vary freely (from 0 to 1.0) on each of the three common, latent A, C, and E factors, and the common pathway model introduces a scaling of the phenotypic variance (i.e., a psychometric factor interpreted as negative emotionality) through which the latent variables act on all three emotions. Findings differed for mother and father reports, as covariance between father report of anger, sadness, and fear was adequately represented by the common pathway model, but the less restrictive independent pathway model was required to explain relations among these emotions according to mother report.

Consistent with past studies of mother report negative emotionality (e.g., Mullineaux et al., 2009; Tackett et al., 2011), we found that a common set of (largely genetic) factors does influence each emotion, supporting the validity of dispositional negative emotionality. However, all three emotions were not equally influenced by this shared A. According to both mother and father reports, genetic influences on sadness were fully shared with anger and fear, but a majority of the variance in fear, and a smaller but still substantial amount of variance in anger, were explained by emotion-specific A factors. Furthermore, we found a small C factor that was shared between mother report anger and sadness but not fear, likely explaining why a common factor was insufficient to represent the genetic and environmental structure of mother report negative emotionality. Thus, we extend past research by suggesting that a less restrictive model enables the detection of shared environmental influences not evident at the broad factor level.

Reporter differences in the heritability and structure of temperament have many potential explanations. We found no evidence of contrast or cooperation effects (which inflate genetic and shared environmental effects, respectively), but the finding of shared environment for mother but not father report may reflect biases leading mothers to rate all children more similarly (Bartels et al., 2003). However, reporter differences also likely reflect true behavioral differences, perhaps due to different contexts in which parents interact with children. Mothers still spend more time in childrearing activities, and are more likely to be primary caregivers (Pleck & Masciadrelli, 2004), and thus may be more familiar with aspects of twins' behavior influenced by the home environment and daily context, potentially increasing shared environmental variance.

Another question is which aspects of the shared environment might influence the development or expression of anger and sadness. Past research in this twin sample finds that the quality of the physical home environment (e.g., safety, cleanliness) moderates the heritability of negative emotionality, with both heritability and total variance higher under adverse conditions (Lemery-Chalfant et al., 2013). This suggests that a safe, structured home environment may constrain the expression of genetic influences on negatively reactive temperament. Furthermore, the emotional climate in the home or the use of similar parenting behaviors with both twins (particularly in frustrating or upsetting situations) may elicit in-the-moment similarities in anger or sadness. To the degree that mothers are parenting in such situations, this is one explanation for the higher shared environmental variance in mother report relative to father report of these emotions.

The finding that fear was largely genetically distinct from anger and sadness regardless of reporter is unsurprising. The developmental literature suggests that observed anger and fear show some stability over time, but are not longitudinally related to each other from infancy to middle childhood (Rothbart et al., 2000), and studies of the physiological correlates of approach and withdrawal emotion and behavior (Fox et al., 2005) also support distinctions between fear and anger. Anger in the current study was more strongly related to sadness than fear for both genetic and shared environmental reasons, but also had moderate emotion-specific A. This A factor may index genetic influences shared with approach but not withdrawal-related aspects of temperament, consistent with research showing that anger shares A and E variance with approach and anticipatory positive affect (Deater-Deckard et al., 2010). This finding is also consistent with a recent study examining the phenotypic structure of observed temperament in preschoolers, which extracted separate factors for dysphoria (anger, sadness, and hostility) and fear-behavioral inhibition (Dyson, Olino, Durbin, Goldsmith, & Klein, 2012). Dyson et al. (2012) suggest that the age at which fear and dysphoria begin to show a structure more similar to adult negative emotionality is an important developmental question. We provide evidence for the continuing independence of fear in middle childhood, even according to parent report.

A substantial literature examines relations between temperament and psychopathology, with evidence for both broad and specific mechanisms (e.g., low effortful control as a risk for multiple disorders vs. anger as a risk for conduct problems), but much of this research has taken place at the level of broad negative emotionality or neuroticism (Nigg, 2006). Considering temperament at the subordinate level may help clarify multiple genetically and environmentally influenced pathways to disorder. For instance, a combination of fear and anger proneness may be important for reactive aggression, to a greater degree than fear or anger alone, whereas low fear but not necessarily high anger may increase risk of instrumental aggression (Nigg, 2006). Furthermore, our finding that genetic variance in fear is less well represented than anger and sadness at the level of broad negative emotionality may be one reason why negative emotionality is often more predictive of externalizing than internalizing symptoms (Rothbart & Bates, 2006), which are more related to fear (Lemery et al., 2002; Oldehinkel et al., 2007). Negative emotionality may capture aspects of reactive distress that are relevant to both externalizing and internalizing (Rothbart & Bates, 2006), but less relevant for withdrawal-specific pathways to internalizing (e.g., via social withdrawal), whereas early behavioral inhibition relatively consistently predicts internalizing problems across childhood and adolescence (Degnan & Fox, 2007). Measurement of both fearfulness or inhibition and proneness to anger and sadness or general distress may be valuable for studies intending to capture multiple pathways from temperament to disorder.

In contrast to parent report, observed anger and sadness were largely distinct, with anger moderately heritable, sadness more influenced by the shared environment, and genetic effects explaining the modest correlation between them. Correlations between observed fear and both observed anger and sadness were too low to allow an examination of genetic and environmental covariance. These modest relations are consistent with observed phenotypic data in preschoolers (e.g., Dyson et al., 2012), but relations between observed anger and sadness in the present study may be due to episode context shared between anger and sadness but not fear. Furthermore, the low correlations between episodes, both within and across emotion, hold implications for the study of temperament using observational measures. The facets of children's emotional reactivity captured by each episode are not the same, and measuring a construct defined by consistency across situations, such as negative emotionality, will likely require multiple and repeated observed episodes. At the same time, children's reactivity within specific episodes may also be meaningful, due to aspects of adjustment captured by responses to unfair sharing, criticism, or other scenarios that may not be tapped by broader measures of negative emotionality.

Limitations

This study is subject to several limitations. First, due to a limited number of observed episodes and relatively low correlations between observed emotions across episodes, aggregating data across multiple episodes was not always possible. Aggregation increases reliability and convergence with parent report (Forman et al., 2003; Rothbart & Bates, 2006). In addition, we derived anger and sadness composites from the same episodes, leading to the possibility that covariance between these emotions was partially or fully due to shared episode context. During the development of the Lab-TAB, episodes intended to tap pure frustration or loss consistently elicited anger in some children and sadness in others. In the current study, we were unable to form anger and sadness composites by aggregating across separate episodes due to low cross-episode correlations, complicating the interpretation of findings for anger and sadness. Anger and sadness were not highly correlated across or within episode (.08–.24), suggesting the relative independence of these emotions even given shared episode context, but these correlations may have been reduced by scoring constraints (i.e., a child cannot be scored as both angry and sad on a single parameter within the same epoch). It will be important for future studies to address these issues using more extensive measurement of observed anger and sadness.

The generalizability of our results to diverse populations is also limited, as participants were largely Caucasian and middle class, and a majority had parents with at least some college education. Although the demographic characteristics of the sample are representative for Wisconsin, heritability is specific to the population under consideration (Plomin et al., 2013). Two other questions specific to twin research are whether findings generalize to singletons and whether environments experienced by MZ twins are more similar than those experienced by DZ twins in a way that is relevant to a trait of interest (i.e., equal environments assumption). Infant twins do not differ from singletons on temperament (Goldsmith & Campos, 1990), and adult twins and singleton siblings do not differ on personality (Johnson, Krueger, Bouchard, & McGue, 2002). Other research also suggests that the equal environments assumption holds for personality and temperament, as greater similarity in childhood experiences appears unrelated to similarity in personality, and differences in infant temperament relate to actual rather than perceived zygosity (Borkenau, Riemann, Angleitner, & Spinath, 2002; Goldsmith et al., 1999).

Finally, we did not model gene–environment interplay, although both gene–environment interaction and gene–environment correlation are likely common (Lemery-Chalfant, 2010). Gene–environment interaction with family-level factors (e.g., SES) may increase A because this interaction increases differences among genetically nonidentical individuals who share the same environment, whereas gene–environment interaction with factors unique to individuals decreases the resemblance of both MZ and DZ twins and may increase E. Thus, while the A factor provides a relatively good estimate of the broad genetic influence on a trait, including gene–environment interplay, small or nonsignificant C in classic twin designs should not be taken as evidence for the irrelevance of family-level influences on child development (Lemery-Chalfant et al., 2013).

Future Directions

It is important to know what aspects of discrete negative emotions may be influenced by a single set of shared genetic factors. One possibility is general reactivity to aversive, threatening, or surprising stimuli, which may manifest as different discrete emotions depending on the context, or a lower physiological threshold of responsivity (Rothbart & Bates, 2006). Genetic influences on self-regulation may also modulate reactive temperament across a range of emotional and behavioral tendencies. Negative emotionality and effortful control are conceptually distinct but empirically negatively correlated in U.S. samples (Rothbart et al., 2001), and may be expected to share genetic variance (e.g., Gagne & Goldsmith, 2011).

Findings also highlight the need for a greater consideration of narrow as well as broad dimensions of temperament, perhaps especially when examining relations between temperament and psychopathology. For instance, irritability (an aversive reaction to stimulation) may share genetic influences with sadness or fear but not frustration (a response to blocked resources or goals), which may be more related to approach and impulsivity. Differentiating between approach-related anger and anger in response to threat or overstimulation may be informative for examining different pathways to risk for externalizing problems, particularly if these aspects of anger are etiologically distinct. Finally, future studies should more closely examine reporter similarities and differences in the genetic and environmental underpinnings of temperament, and potential differences in prediction of child psychopathology and adaption, as mothers, fathers, and trained laboratory observers may each be capable of detecting, or even eliciting, distinct aspects of children's emotional responses. In addition, although child-report versions of the CBQ have not yet been developed, the development of a corresponding child-report measure of emotion and comparison with parent report is one important avenue for future research.

Taken as a whole, the primary implication of our findings is that the genetic and environmental underpinnings of temperament may differ greatly depending on the level of analysis and the context. Findings support the use of broad negative emotionality, but there was evidence for heritable aspects of anger and fear that cannot be accounted for by a shared set of genetic influences, and shared environmental influences on mother report anger and sadness that were not shared with fear. Thus, what is common to all three emotions is meaningful but unable to encompass the full extent of systematic, biologically based individual differences in discrete emotions. Fear, in particular, is not well captured by broad measures of negative emotionality. In light of these distinctions, future research should consider the unique contributions of narrow facets of temperament to children's psychological and socioemotional functioning, both independently and in interaction with other aspects of children's temperament and environmental context.

References

Akaike, H. (1987). Factor analysis and AIC. *Psykometrika, 52*, 317–332. doi:10.1007/BF02294359

Bartels, M., Hudziak, J. J., Boomsma, D. I., Rietveld, M. J., van Beijsterveldt, T. C., & van den Oord, E. J. (2003). A study of parent ratings of internalizing and externalizing problem behavior in 12-year-old twins. *Journal of the American Academy of Child and Adolescent Psychiatry, 42*, 1351–1359. doi:10.1097/01.CHI.0000085755.71002.5d

Boker, S. M., Neale, M. C., Maes, H. H., Wilde, M. J., Spiegel, M., Brick, T. R., . . . Fox, J. (2011). OpenMx: An open source extended structural equation modeling framework. *Psychometrika, 76*, 306–317. doi:10.1007/s11336-010-9200-6

Borkenau, P., Riemann, R., Angleitner, A., & Spinath, F. M. (2002). Similarity of childhood experiences and personality resemblance in monozygotic and dizygotic twins: A test of the equal environments assumption. *Personality and Individual Differences, 33*, 261–269. doi:10.1016/S0191-8869(01)00150-7

Carver, C. S., & Harmon-Jones, E. (2009). Anger is an approach-related affect: Evidence and implications. *Psychological Bulletin, 135*, 183–204. doi:10.1037/a0013965

Deater-Deckard, K., Beekman, C., Wang, Z., Kim, J., Petrill, S. A., Thompson, L. A., & DeThorne, L. S. (2010). Approach/positive anticipation, frustration/anger, and overt aggression in childhood. *Journal of Personality, 78*, 991–1010. doi:10.1111/j.1467-6494.2010.00640.x

Deater-Deckard, K., Petrill, S. A., & Thompson, L. A. (2007). Anger/frustration, task persistence, and conduct problems in childhood: A behavioral genetic analysis. *Journal of Childhood Psychology and Psychiatry, 48*, 80–87. doi:10.1111/j.1469-7610.2006.01653.x

Degnan, K., & Fox, N. (2007). Behavioral inhibition and anxiety disorders: Multiple levels of a resilience process. *Development and Psychopathology, 19*, 729–746. doi:10.1017/s0954579407000363

De Los Reyes, A., Henry, D., Tolan, P., & Wakschlag, L. (2009). Linking informant discrepancies to observed variations in young children's disruptive behavior. *Journal of Abnormal Child Psychology, 37*, 637–652. doi:10.1007/s10802-009-9307

DiLalla, L. F., Kagan, J., & Reznick, J. S. (1994). Genetic etiology of behavioral inhibition among 2-year-old children. *Infant Behavior and Development, 17*, 405–412. doi:10.1016/0163-6383(94)90032-9

Dyson, M. W., Olino, T. M., Durbin, E. C., Goldsmith, H. H., & Klein, D. N. (2012). The structure of temperament in preschoolers: A two-stage factor analytic approach. *Emotion, 12*, 44–57. doi:10.1037/a0025023

Eisenberg, N., Valiente, C., Spinrad, T. L., Cumberland, A., Liew, J., Reiser, M., & Losoya, S. H. (2009). Longitudinal relations of children's effortful control, impulsivity, and negative emotionality to their externalizing, internalizing, and co-occurring behavior problems. *Developmental Psychology, 45,* 988–1008. doi:10.1037/a0016213

Emde, R. N., Plomin, R., Robinson, J., Corley, R., DeFries, J., Fulker, D. W., . . . Zahn-Waxler, C. (1992). Temperament, emotion, and cognition at fourteen months: The MacArthur Longitudinal Twin Study. *Child Development, 63,* 1437–1455. doi:10.1111/j.1467-8624.1992.tb01706.x

Emde, R. N., Robinson, J. L., Corley, R. P., Nikkari, D., & Zahn-Waxler, C. (2001). Reactions to restraint and anger-related expressions during the second year. In R. N. Emde & J. K. Hewitt (Eds.), *Infancy to early childhood: Genetic and environmental influences on developmental change* (pp. 127–140). New York, NY: Oxford University Press.

Forman, D. R., O'Hara, M. W., Larsen, K., Coy, K. C., Gorman, L. L., & Stuart, S. (2003). Infant emotionality: Observational methods and the validity of maternal reports. *Infancy, 4,* 541–565. doi:10.1207/S15327078IN0404_08

Fox, N., Henderson, H., Marshall, P., Nichols, K., & Ghera, M. (2005). Behavioral inhibition: Linking biology and behavior within a developmental framework. *Annual Review of Psychology, 56,* 235–262. doi:10.1146/annurev.psych.55.090902.141532

Gagne, J. R., & Goldsmith, H. H. (2011). A longitudinal analysis of anger and inhibitory control in twins from 12 to 36 months of age. *Developmental Science, 14,* 112–124. doi:10.1111/j.1467-7687.2010.00969.x

Gagne, J. R., & Saudino, K. J. (2010). Wait for it! A twin study of inhibitory control in early childhood. *Behavior Genetics, 40,* 327–337. doi:10.1007/s10519-009-9316-6

Gagne, J. R., Van Hulle, C. A., Aksan, N., Essex, M. J., & Goldsmith, H. H. (2011). Deriving childhood temperament measures from emotion-eliciting behavioral episodes: Scale construction and initial validation. *Psychological Assessment, 23,* 337–353. doi:10.1037/a0021746

Gagne, J. R., Vendlinski, M. K., & Goldsmith, H. H. (2009). The genetics of childhood temperament. In K. Yong-Kyu (Ed.), *The handbook of behavior genetics* (pp. 251–267). New York, NY: Springer. doi:10.1007/978-0-387-76727-7_18

Goldsmith, H. H. (1991). A zygosity questionnaire for young twins: A research note. *Behavior Genetics, 21,* 257–269. doi:10.1007/BF01065819

Goldsmith, H. H., Buss, K. A., & Lemery, K. S. (1997). Toddler and childhood temperament: Expanded content, stronger genetic evidence, new evidence for the importance of environment. *Developmental Psychology, 33,* 891–905. doi:10.1037//0012-1649.33.6.891

Goldsmith, H. H., & Campos, J. J. (1990). The structure of temperamental fear and pleasure in infants: A psychometric perspective. *Child Development, 61,* 1944–1964. doi:10.1111/j.1467-8624.1990.tb03577.x

Goldsmith, H. H., & Lemery, K. L. (2000). Linking temperamental fearfulness and anxiety symptoms: A behavior-genetic perspective. *Biological Psychiatry, 48,* 1199–1209. doi:10.1016/S0006-3223(00)01003-9

Goldsmith, H. H., Lemery, K. S., Buss, K. A., & Campos, J. J. (1999). Genetic analyses of focal aspects of infant temperament. *Developmental Psychology, 35,* 972–985. doi:10.1037//0012-1649.35.4.972

Goldsmith, H. H., Reilly, H. H., Lemery, K. L., Longley, S., & Prescott, A. (2001). *The Laboratory Temperament Assessment Battery: Middle childhood version.* Unpublished manuscript, University of Wisconsin-Madison.

Izard, C. E. (2007). Basic emotions, natural kinds, emotion schemas, and a new paradigm. *Perspectives on Psychological Science, 2,* 260–280. doi:10.1111/j.1745-6916.2007.00044.x

Johnson, W., Krueger, R. F., Bouchard, T. J., & McGue, M. (2002). The personalities of twins: Just ordinary folks. *Twin Research, 5,* 125–131. doi:10.1375/1369052022992

Kagan, J., & Fox, N. A. (2006). Biology, culture and temperamental biases. In W. Damon & N. Eisenberg (Eds.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (pp. 167–223). New York, NY: Wiley.

Karevold, E., Coplan, R., Stoolmiller, M., & Mathiesen, K. (2012). A longitudinal study of the links between temperamental shyness, activity, and trajectories of internalising problems from infancy to middle childhood. *Australian Journal of Psychology, 63,* 36–43. doi:10.1111/j.1742-9536.2011.00005.x

Lemery, K. S., Essex, M. J., & Smider, N. A. (2002). Revealing the relation between temperament and behavior problem symptoms by eliminating measurement confounding: Expert ratings and factor analyses. *Child Development, 73,* 867–882. doi:10.1111/1467-8624.00444

Lemery-Chalfant, K. (2006). Wisconsin Twin Panel: Current directions and findings. *Twin Research and Human Genetics, 9,* 1030–1037. doi:10.1375/twin.9.6.1030

Lemery-Chalfant, K. L. (2010). Genes and environments: How they work together to promote resilience. In J. Reich, A. J. Zautra, & J. S. Hall (Eds.), *Handbook of adult resilience* (pp. 55–78). New York, NY: Guilford.

Lemery-Chalfant, K., Clifford, S., & Swann, G. (in press). Temperament and child psychopathology: Specificity in shared genetic effects. In Y. Kim (Series Ed.) & K. Saudino & J. Ganiban (Vol. Eds.), *Behavior genetics book series: Behavior genetics of temperament and personality.* New York, NY: Springer.

Lemery-Chalfant, K., Kao, K., Swann, G., & Goldsmith, H. (2013). Childhood temperament: Passive gene-environment correlation, gene-environment interaction, and the hidden importance of the family environment. *Development and Psychopathology, 25,* 51–63. doi:10.1017/S0954579412000892

Martin, N. G., & Eaves, L. J. (1977). The genetical analysis of covariance structure. *Heredity, 38,* 79–95. doi:10.1038/hdy.1977.9

McArdle, J. J., & Goldsmith, H. H. (1990). Alternative common factor models for multivariate biometric analyses. *Behavior Genetics, 20,* 569–608. doi:10.1007/BF01065873

McGue, M., & Bouchard, T. J. (1984). Adjustment of twin data for the effects of age and sex. *Behavior Genetics, 14,* 325–343. doi:10.1007/BF01080045

McGuire, S., Clifford, J., Fink, J., Basho, S., & McDonnell, A. M. (2003). Children's reactions to the unfamiliar in middle childhood and adolescence: An observational twin/sibling study. *Personality and Individual Differences, 35,* 339–354. doi:10.1016/S0191-8869(02)00194-0

Mullineaux, P., Deater-Deckard, K., Petrill, S., Thompson, L., & DeThorne, L. (2009). Temperament in middle childhood: A behavioral genetic analysis of fathers' and mothers' reports. *Journal of Research in Personality, 43,* 737–746. doi:10.1016/j.jrp.2009.04.008

Murphy, F. C., Nimmo-Smith, I., & Lawrence, A. D. (2003). Functional neuroanatomy of emotion: A meta-analysis. *Cognitive, Affective, & Behavioral Neuroscience, 3,* 207–233. doi:10.3758/CABN.3.3.207

Muthén, B., & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology, 38,* 171–189. doi:10.1111/j.2044-8317.1985.tb00832.x

Muthén, L. K., & Muthén, B. O. (1998–2011). *Mplus user's guide* (6th ed.). Los Angeles, CA: Author.

Neale, M. C., & Cardon, L. R. (1992). *Methodology for genetic studies of twins and families*. Dordrecht, Netherlands: Kluwer Academic.

Nigg, J. T. (2006). Temperament and developmental psychopathology. *Journal of Child Psychology and Psychiatry, 47,* 395–422. doi:10.1111/j.1469-7610.2006.01612.x

Oldehinkel, A. J., Hartman, C. A., Ferdinand, R. F., Verhulst, F. C., & Ormel, J. (2007). Effortful control as moderator of the association between negative emotionality and adolescents' mental health problems. *Development and Psychopathology, 19,* 523–539. doi:10.1017/S0954579407070253

OpenMx Development Team. (2011). *OpenMx official documentation.* Retrieved from http://openmx.psyc.virginia.edu/docs/openmx/latest/_static/rdoc/mxCI.html

Pleck, J. H., & Masciadrelli, B. P. (2004). Parental involvement in U.S. residential fathers: Levels, sources, and consequences. In M. E. Lamb (Ed.), *The role of the father in child development* (4th ed., pp. 222–271). New York, NY: Wiley.

Plomin, R., DeFries, J. C., Knopik, V. S., & Neiderhiser, J. M. (2013). *Behavioral genetics* (6th ed.). New York, NY: Worth.

Price, T. S., Freeman, B., Craig, I. W., Petrill, S. A., Ebersole, L., & Plomin, R. (2000). Infant zygosity can be assigned by parental report questionnaire data. *Twin Research, 3,* 129–133. doi:10.1375/136905200320565391

Rhee, S. H., Corley, R. P., Friedman, N. P., Hewitt, J. K., Hink, L. K., Johnson, D. P., . . . Young, S. E. (2012). The etiology of observed negative emotionality from 14 to 24 months. *Frontiers in Genetics, 3.* doi:10.3389/fgene.2012.00009

Robinson, J., Kagan, J., Reznick, J. S., & Corley, R. (1992). The heritability of inhibited and uninhibited behavior: A twin study. *Developmental Psychology, 28,* 1030–1037. doi:10.1037//0012-1649.28.6.1030

Rothbart, M. K., Ahadi, S. A., Hershey, K., & Fisher, P. (2001). Investigations of temperament at 3 to 7 years: The children's behavior questionnaire. *Child Development, 72,* 1394–1408. doi:10.1111/1467-8624.00355

Rothbart, M. K., & Bates, J. E. (2006). Temperament. In W. Damon & N. Eisenberg (Eds.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (pp. 105–176). New York, NY: Wiley. doi:10.1002/9780470147658.chpsy0303

Rothbart, M. K., Evans, D. E., & Ahadi, S. A. (2000). Temperament and personality: Origins and outcomes. *Journal of Personality and Social Psychology, 78,* 122–135. doi:10.1037//0022-3514.78.1.122

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39,* 1161–1178. doi:10.1037/h0077714

Saudino, K. J. (2003). Parent ratings of infant temperament lessons from twin studies. *Infant Behavior and Development, 26,* 100–107. doi:10.1016/S0163-6383(02)0017-6

Saudino, K. J. (2005). Behavioral genetics and child temperament. *Journal of Developmental and Behavioral Pediatrics, 26,* 214–223. doi:10.1097/00004703-200506000-00010

Schwartz, C. E., Snidman, N., & Kagan, J. (1996). Early childhood temperament as a determinant of externalising behaviour in adolescence. *Development and Psychopathology, 8,* 527–537. doi:10.1017/S0954579400007252

Singh, A. L., & Waldman, I. D. (2010). The etiology of associations between negative emotionality and childhood externalizing disorders. *Journal of Abnormal Psychology, 119,* 376–388. doi:10.1037/a0019342

Tackett, J. L., Waldman, I. D., Van Hulle, C. A., & Lahey, B. (2011). Shared genetic influences on negative emotionality and major depression/conduct disorder comorbidity. *Journal of the American Academy of Child and Adolescent Psychiatry, 50,* 818–827. doi:10.1016/j.jaac.2011.05.007

Tooby, J., & Cosmides, L. (2008). The evolutionary psychology of emotions and their relationship to internal regulatory variables. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of emotions* (3rd ed., pp. 114–137). New York, NY: Guilford.

Vytal, K., & Hamann, S. (2010). Neuroimaging support for discrete neural correlates of basic emotions: A voxel-based meta-analysis. *Journal of Cognitive Neuroscience, 22,* 2864–2885. doi:10.1162/jocn.2009.21366

Child Development, September/October 2015, Volume 86, Number 5, Pages 1557–1570

Reciprocal Relations Between Student–Teacher Relationship and Children's Behavioral Problems: Moderation by Child-Care Group Size

Věra Skalická
NTNU Social Research

Jay Belsky
University of California, Davis

Frode Stenseng
NTNU Social Research

Lars Wichstrøm
NTNU Social Research and NTNU

In this Norwegian study, bidirectional relations between children's behavior problems and child–teacher conflict and closeness were examined, and the possibility of moderation of these associations by child-care group size was tested. Eight hundred and nineteen 4-year-old children were followed up in first grade. Results revealed reciprocal effects linking child–teacher conflict and behavior problems. Effects of child–teacher closeness on later behavior problems were moderated by group size: For children in small groups only (i.e., ≤ 15 children), greater closeness predicted reduced behavior problems in first grade. In consequence, stability of behavior problems was greater in larger than in smaller groups. Results are discussed in light of regulatory mechanisms and social learning theory, with possible implications for organization of child care.

Children with externalizing behavior problems tend to be hyperactive, impulsive, and aggressive, performing less well in school than other children (O'Connor, Dearing, & Collins, 2011). Externalizing behavior problems are evident during the preschool years, tend to persist into school age, and can extend into adulthood, with possible results such as low educational achievement, unemployment, and criminality (O'Connor et al., 2011).

In addition to established child and family predictors of behavioral problems, evidence indicates that child-care experience also matters (Belsky et al., 2007; National Institute of Child Health and Human Development [NICHD] Early Child Care Research Network, 2003). Prior research has examined whether and how proximate markers of child-care quality (e.g., caregiver sensitivity, cognitive stimulation) as well as more structural features (e.g., type, group size, time spent in child care) may influence children's development (e.g., McCartney et al., 2010; NICHD Early Child Care Research Network, 2006). Of particular importance for this report is evidence that the quality of relationships that children experience with their caregivers is predictive of children's behavior problems (Sabol & Pianta, 2012).

Consistent with ecological (Bronfenbrenner & Morris, 1998) and transactional (Sameroff & Mackenzie, 2003) theoretical frameworks that underscore the ongoing reciprocal relations between children and their surroundings, preschool relationships with teachers not only influence children's development but are also expected to be affected by characteristics of children (Birch & Ladd, 1998; Sutherland & Oswald, 2005). Of note, however, is that evidence of such reciprocal effects in the case of children's behavior and the student–teacher relationship (STR) is limited (Nurmi, 2012). One of the few relevant studies found that more child aggression at the beginning of the kindergarten year predicted increased teacher–child conflict, which in turn forecast an increase in child aggression by the end of the school year (Doumen et al., 2008). In the present research, we sought to extend prevailing knowledge by examining bidirectional processes involving two aspects of STR quality—conflict and closeness—and children's externalizing behavior; additionally, we sought to evaluate whether the assumed reciprocal relations persist across the transition from preschool to school.

Furthermore, prior research provides evidence that effects of teacher–child relationships can be

---

This research was supported by the Norwegian Research Council (203249).

Correspondence concerning this article should be addressed to Věra Skalická, NTNU Social Research, Dragvoll alle 38B, 7491 Trondheim, Norway. Electronic mail may be sent to vera.skalicka@samfunn.ntnu.no.

© 2015 The Authors
Child Development © 2015 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2015/8605-0017
DOI: 10.1111/cdev.12400

moderated by children's characteristics (e.g., gender: Baker, 2006; race: Murray, Waas, & Murray, 2008; externalizing behavior: Baker, 2006; Silver, Measelle, Armstrong, & Essex, 2005; effortful control: Liew, Chen, & Hughes, 2010). Classroom context has also been regarded as a potential moderator of child effects on their later development (Buyse, Verschueren, Doumen, Van Damme, & Maes, 2008). In this regard, the research reported herein is the first to test the proposition that preschool group size can moderate the putative influence of preschool teacher–child relationship or child behavior on children's future functioning. Therefore, the second goal of the present study was to investigate whether the preschool context, operationalized here in terms of group size, can alter (a) discerned effects of teacher–child relationship on the development of behavior problems, as well as (b) anticipated effects of children's behavior on subsequent behavior and quality of STR.

STR and Problem Behavior—Reciprocal Relations

Pianta (1999) theorized, consistent with attachment theory, that the STR can function as a secure base to aid and abet children's adjustment to school settings and, thereby, their positive development. For such a sense of felt security to develop—and with it the capacity to self-regulate emotion and behavior—the teacher should accurately perceive the child's needs and desires and behave in a supportive and caring manner (Pianta, 1999; Pianta, Hamre, & Stuhlman, 2003). In consequence, children should learn or continue to use adaptive strategies when engaging their social environment (Silver et al., 2005). Thus, high-quality STRs, characterized by high levels of closeness and low levels of conflict, are presumed to promote children's self-regulatory capacities (Pianta, 1999). Positive STRs might help children to develop—or maintain—positive internal working models of self and others, encourage them to seek support from others, while enabling them to manage their emotions in a socially acceptable manner (O'Connor et al., 2011). As a result of such supportive relationship experiences, children should be motivated to engage in appropriate behavior (Hamre & Pianta, 2001). Not surprisingly, then, positive STRs have been found to protect against the development of behavior problems (Buyse et al., 2008; Hughes & Cavell, 1999; O'Connor et al., 2011; Silver et al., 2005). In contrast, high levels of conflict and/or low levels of closeness with the teacher predict increased aggressive behavior over time (Birch & Ladd, 1997; Pianta & Stuhlman, 2004; Silver et al., 2005). Low-quality STRs may lead teachers to be less sensitive, provide less support, and try to control children's behavior (Hamre & Pianta, 2001), which might result in cycles of coercive interaction (Sutherland & Oswald, 2005), thereby fostering the development of maladaptive strategies of emotion regulation or conflict resolution.

Effects of STRs in preschool tend to endure over several years (Hamre & Pianta, 2001), thereby influencing future relationships with teachers in school (Howes, Phillipsen, & Peisner-Feinberg, 2000; Jerome, Hamre, & Pianta, 2009; O'Connor & McCartney, 2006). Nevertheless, it must be appreciated—as the concept of coercive cycles just mentioned clearly implies—that influence processes are bidirectional (Sameroff & Mackenzie, 2003; Sutherland & Oswald, 2005). Thus, greater behavior problems on the part of the child predict more problematic STRs (Nurmi, 2012), with the reverse being true of lower levels of behavior problems (O'Connor, 2010). Indeed, two studies chronicled reciprocal effects linking children's externalizing behavior and teacher–child conflict in preschool (Zhang & Sun, 2011) and in kindergarten (Doumen et al., 2008). What has not yet been established is whether discerned reciprocal effects also apply to teacher–child closeness and/or persist across the transition from preschool to school, the developmental focus of the current inquiry.

In addition to child characteristics, the reciprocally dynamic STR can be influenced by factors and forces beyond the school setting, most notably the family (O'Connor, 2010). This would be consistent with Bronfenbrenner's (1977) notion of the "mesosystem" or interrelation of immediate settings in which the child spends substantial time. This implies that the background family context should be taken into account when examining children's experiences and functioning in preschool and elementary school.

Furthermore, whereas previous research has focused almost exclusively on teacher-reported problem behavior (e.g., Doumen et al., 2008; Silver et al., 2005), in the present study we extend inquiry by examining whether effects of STR might translate on general aspects of children's behavior, not only in the school, but also in other social arenas, as seen through the eyes of the child's parents.

The Moderating Role of Preschool Group Size

Based on the notion that reciprocal, cross-time relations between teacher–child relationships and behavior problems—the focus of interest here—are probabilistic and can vary as a function of the context in which these relationships operate (Sameroff, 1975; Sameroff & Mackenzie, 2003), we examine the moderating role of preschool classroom context. Much research underscores the importance of classroom relational climate or classroom aggression levels, as predictors of the STR (Mantzicopoulos, 2005) or as predictors of changes in student aggression (Thomas, Bierman, Powers, & the Conduct Problems Prevention Research, 2011). Such evidence raises the possibility that classroom contextual factors could affect the extent and/or nature of cross-time links between caregiver/teacher–child relationships and behavior problems. However, evidence of such contextual moderation is limited to only one study showing that the effect of children's externalizing problems on subsequent conflict with teachers proved most pronounced when children were in classrooms marked by high levels of problem behavior (Buyse et al., 2008). What these investigators did not consider, however, was whether a reciprocal process is operative as well, and thus whether this caregiver/teacher–child relationship effects on child behavior are also moderated by classroom context. One of the goals of the current study is to address this empirical lacuna and, thereby, extend inquiry in this area. This is done by focusing on group size as a contextual moderator of reciprocal, cross-time links between caregiver/teacher–child relationships and problem behavior.

We focus on group size, because it has been linked with children's functioning. More specifically, children who spend a greater proportion of their time in child care in a large group of peers engage in more externalizing behavior than other children (e.g., McCartney et al., 2010). This may be because children in small groups spend less time interacting with peers and more time interacting with adults than they do in larger groups (e.g., Dunn, 1993). Furthermore, in smaller groups, the teacher might engage in more direct interaction with the individual child (Bourke, 1986) compared to children in larger groups, who gain less individual attention from adults (Blatchford, 2003; Pianta et al., 2003). Structural characteristics of preschool groups in terms of their size might not only directly affect the quality of child–teacher interactions, such as the teacher's sensitivity (Pianta et al., 2003), but they might also possibly exacerbate effects of already established behaviors and relationships on later functioning. Based on these observations, we expect that experience in *smaller* groups will enhance the influence of child–teacher relationship quality on subsequent child behavior. This social-learning-theory-based hypothesis (Bandura, 1971) stems from the view that in small groups, the teacher will be a greater source of influence on children than in large groups. We further expect that experience in *larger* groups will strengthen the links between children's problem behavior and later children's functioning due to learning of antisocial behavior from peers.

Group Size as a Moderator of the STR—Children's Behavior Link

With regard to the link between STR and later children's functioning, we assume that children in small groups will be more likely to learn to regulate their emotions and behavior when they experience closer and less conflicted relationships with teachers than when they spend more time in large groups. What this hypothesis implies, of course, is that small groups should enhance the influence of the STR in a for-better-and-for-worse manner: When children are in small groups and child–teacher relationships are more positive, children should manifest fewer problems over time, but when such relationships are more conflicted in small groups, children should develop more problems than would be the case were they in larger groups.

In other words, we predict that when group sizes are smaller, teacher–child closeness will exert a greater protective effect on future behavior problems than when they are larger. This expectation is based on the view that in smaller classes in which teachers can spend more time with and get to know children better, the effects of closeness will be more pronounced in preventing the growth of behavior problems. However, when teachers experiencing conflicted relationships with students are forced to attend to a child more frequently—as is presumably the case in smaller groups—they might participate in more coercive interactions, which might in turn result in the fostering of problem behavior. These processes, we suspect, would be less likely to operate in larger groups where individualized attention may be less frequent.

We contend that predictions advanced through this point should receive empirical support even when the adult–child ratio remains constant as group size varies—as it typically does in the Norwegian context where the current research was carried out (i.e., six children per adult). We base our reasoning on the view that smaller groups are more likely to remain stable throughout the day, so that the child has the opportunity to interact with the same teacher over hours, days, weeks, and months. In large groups, efforts to create smaller subgroups—so that special activities can be pursued—often result in children dealing with different teachers throughout the day. And in situations in which large groups remain undivided, the complexity of interactions and interactants (child–child; teacher–teacher; child–teachers) likely results in less opportunity for children to benefit from established close relationships with an individual teacher to the same extent as children from smaller groups. Thus, even though the teacher–child ratio in larger groups remains the same as in smaller groups, teachers have to divide their attention between a larger number of children. By the same token, the hypothesized negative effects of a conflicted relationship with one teacher on later children's functioning will be less pronounced in larger groups compared to smaller groups. This line of argument is supported by evidence from qualitative research conducted in Norway, revealing that teachers in larger centers that typically have larger groups report that it is more difficult for them than for teachers in smaller centers who manage smaller groups to pay attention to individual children (Alvestad et al., 2014; Seland, 2011).

Group Size as a Moderator of Children's Behavior—Later Functioning Link

Furthermore, it is possible that group size can moderate the stability of problem behavior. We hypothesize that when children are exposed to more peers in large preschool groups that negative peer influence will be greater; thus, preschool-age problem behavior will predict greater problem behavior on the part of first grade-age children than would be the case were groups smaller. This prediction is based on the social-learning view that in larger groups, children often learn antisocial behavior from peers (Laird, Jordan, Dodge, Pettit, & Bates, 2001). There is also empirical evidence supporting this peer learning mechanism in a school setting. Barth, Dunlap, Dane, Lochman, and Wells (2004) found that children already showing behavior problems in fourth grade displayed a greater increase in problem behavior when placed in highly aggressive fifth-grade classroom environments than when placed in less aggressive classrooms. Therefore, it is possible that children already displaying behavior problems might be more vulnerable to the influence of large group size and will thus be more prone to manifest such problem behavior 2 years later.

There is further evidence consistent with the proposition just advanced that large groups may promote problem behavior. An American study by McCartney et al. (2010) showed it was principally when children were in large groups—which proved to be the norm—that greater exposure to child care predicted more externalizing problems. Experience in large peer groups has also been linked to higher stress levels and behavioral and emotional dysregulation (Fabes, Hanish, & Martin, 2003), which might in turn lead to increased problem behavior (Legendre, 2003; Vermeer & van IJzendoorn, 2006).

Based on the aforementioned findings, we argue that this negative peer-learning process will prove more operative in larger groups—though we do not measure this process directly, but only examine its hypothesized effects. We argue that larger groups have a greater probability of containing at least some children with behavior problems. Children in larger groups may thus have more opportunity to be disturbed by or interact with noncompliant children or children with problem behavior than in smaller groups. In other words, contact with larger number of peers increases the probability of observing and potentially reproducing problem behavior, and hence leading to increased stability of behavior problems.

According to Clarke-Stewart (1989), children do not learn to follow social rules or to resolve conflicts without resorting to aggression unless guided by their caregivers. However, larger groups might put extra strain on teachers, thus leaving less opportunity for providing emotional support and appropriate response to children's problem behavior, which might in turn lead to even more problem behavior—even in the face of common teacher–child ratios across smaller and larger classrooms. Therefore, we argue that this additional explanation in terms of less positive teacher-directed regulation of behavior might add to the effects of more negative peer influence, and thereby increases the stability of behavioral problems among children in larger groups.

So far, we have hypothesized that in larger groups compared to smaller groups, the stability of problem behavior will be greater than would be the case were children in smaller groups. Nonetheless, based on the bidirectional perspective, it is possible that group size can also moderate the effect of children's problem behavior on later student–teacher conflict. However, we did not formulate a specific hypothesis regarding this relation and our investigation of this interaction was only exploratory.

Norwegian Child-Care System

In order to familiarize readers with child care in Norway, which has a system quite different from that of the United States, the locale where most of the work cited in this report was conducted, we provide a brief description of its most important regulatory conditions. Norwegian welfare state provides employees with parental leave for 10.5 or 13 months with, respectively, 100% or 80% salary replacement and universal, heavily subsidized access to child care, to which every toddler is entitled upon turning 1 year of age. Consequently, 80% of children enroll in a day-care facility during their 2nd year of life (Statistics Norway, 2013a).

Many aspects of Norwegian child care are regulated by law, including the preschool curriculum, staff training and adult–child ratio. As a result, the quality of Norwegian child care is quite homogeneous and in accord with international standards (UNICEF, 2008). Indeed, Norway achieved 8 of a maximum 10 benchmarks; the United States achieved 3. The Norwegian system prescribes that the adult–child ratio cannot be greater than one caregiver for every six children age 3 or older, and this rule is widely followed (Vassenden, Thygesen, Bayer, Alvestad, & Abrahamsen, 2011). However, Norway lags behind several highly developed countries when it comes to staff education, because it does not meet the suggested criteria that 80% of all child-care staff must be trained (UNICEF, 2008). There must also be at least one teacher—with minimum of 3 years of preschool pedagogical education at college level—per 14–18 children age 3 or older, and a child-care center pedagogical leader must have the same qualifications (Ministry of Education, 2012). However, in 2006, 8% of child-care staff who were supposed to have such education did not meet these requirements (Kjelvik, 2012).

Day care is mainly provided in child-care centers and to a smaller extent in group care in private homes (3.4% in 2008; Norwegian Directorate for Education and Training, 2012). The maximum monthly fee that parents pay—about $ 380—is legally stipulated (Statistics Norway, 2013b). There are virtually no differences between various types of child care, except that in private-home child care the staff is slightly less educated and must be occasionally supervised. Importantly, *size of the peer group is not regulated and can vary between different arrangements* (as long as the standard ratio is maintained). In the current study, we thus aim to investigate the importance of the size of the child-care group when it comes to understanding the cross-time interrelation of STRs and children's behavior problems. Given the presence of ratio- but not group-size-related regulations in the Norwegian child-care system, there will be clear translational implications should the hypotheses advanced receive empirical support.

Method

Participants and Procedures

The Trondheim Early Secure Study comprises participants from two birth cohorts of children born in 2003 and 2004 and their parents living in the city of Trondheim, Norway. Only a brief outline is provided here, as details about the procedure and recruitment have been presented previously (Wichstrøm et al., 2012). The Strengths and Difficulties Questionnaire (SDQ) 4–16 version (Goodman, 1997), together with an invitation letter, were mailed to parents ($N = 3{,}456$). Completed SDQs were brought to a subsequent community health checkup; all 4-year-olds in Norway are encouraged to attend the well-child clinic and 3,358 did so. Parents with inadequate proficiency in Norwegian were excluded ($N = 176$). The health nurse missed asking 166 parents about their interest in participating in the study. At the well-child clinic eligible parents ($N = 3{,}016$) were informed about the study using procedures approved by the Regional Committee for Medical and Health Research Ethics. Written consent was obtained from parents of 2,475 children (71.6% of all contacted).

In order to recruit a study sample that oversampled for children at risk for developing problems, children's SDQ total difficulties scores were divided into four strata. Using a random number generator, defined proportions of children in each stratum, ranging from few to many problems (i.e., 0.37, 0.48, 0.70, 0.89) were drawn to participate in a further study ($N = 1{,}250$). Of the 1,250 parents invited to participate, 992 (77.9%) parents appeared at the university for further study at Time 1 (T1), where parents provided information about child-care history; child behavior, child, and family factors; and children's language comprehension ($N = 935$) was examined. Dropout rate did not vary by SDQ strata, $\chi^2 = 5.70$ (3), $p = .13$, or gender, $\chi^2 = 0.23$ (1), $p = .63$. Eight hundred and nineteen parents participated in a follow-up assessment 2 years later at Time 2 (T2), when the child had started in first grade. Almost as many girls (49.5%) as boys (50.5%) participated at T2. Nearly all parents who participated in the study were the child's biological parent (98.2%), and a large majority of the participating parents were women (84.5%). Both mothers (96.4%) and fathers (94.8%) were mainly of Norwegian ethnicity. The majority of parents was married (54.8%) or had lived together for more than 6 months (34.3%). With regard to marital status, 8.5% were divorced or separated, 0.3% widowed, 1.2% had lived together for < 6 months, and 0.9% of the parents had never lived together.

The parents consented to having their child-care provider and/or teacher complete questionnaires regarding teacher–child relationship quality, which were sent to day-care centers at T1 and to primary schools at T2. Day-care centers and schools were requested to select the teacher who knew the child best to respond to the questionnaire. Response rates among teachers were 90.6% at T1 and 92.2% at T2. Preschool teachers had known the child for an average of 13 months whereas school teachers had known the child for an average of 6 months.

Measures

Externalizing Problem Behavior

The externalizing scale of the parent version of the Child Behavior Checklist (CBCL) was used to assess externalizing problems at T1 and T2 (Achenbach & Rescorla, 2000). At T1 the 1.5–5 year version was applied (25 items) whereas at T2 the 6–18 year version (35 items) was administered. For each item, the parent is asked to determine how well the item describes the child now or within the past 6 months: 0 = *not true*, 1 = *somewhat or sometimes true*, 2 = *very true or often true*. Higher scores indicate more problems. The CBCL has excellent concurrent and predictive validity and is the most widely used screening instrument for behavior problems in children (O'Connor, 2010). We preferred parent reports of CBCL to teacher reports in order to minimize the effects of shared rater variance if teachers were to rate behavior problems and child–teacher relationships. Cronbach's alphas were .89 and .88 at T1 and T2, respectively.

Student–Teacher Relationship

The conflict and closeness subscales of the Student-Teach Relationship Scale (Pianta, 2001) were used to measure child–teacher relationships at preschool and at first grade on a 5-point Likert-type scale ranging from 1= *definitely does not apply* to 5= *definitely applies*. The conflict subscale (12 items) provides teacher-perceived negativity within the relationship with the child, while the closeness subscale (11 items) assesses whether the teacher perceives the relationship to be warm, affectionate, and including open communication (Jerome et al., 2009). Cronbach's alphas were .77 and .82 for conflict, and .65 and .70 for closeness at T1 and T2, respectively.

Child-Care Group Size

Child care was defined as regularly scheduled care, outside of the child's home, provided by a nonrelative to three or more children. Parents reported how many other children were present in the child's child-care group at age 3–4 years. Groups that were smaller than 12 children represented only 7.3% of the sample. Based on the distribution of group size, we assigned children to small groups (up to 15 children) and large groups (more than 15 children). We arrived at cutoff point ≤ 15 because it represents the lowest tertile, but the results were robust for (small) group size up to 17 children. While the issue of group size illuminated in the work by McCartney et al. (2010) informed our study, we have not set out to use the American specific group size cutoffs, but rather aimed for group size cutoffs informative to the Norwegian context. We also tested whether there were any differences regarding strength of moderation of group size on STR and behavior between children from groups counting 16–20 children versus more than 20 children (second vs. third tertile); however, these groups did not differ significantly.

Child and Family Covariates

The associations between quality of the STR and levels of problem behavior (and stability of behavior over time) might be influenced by a variety of potentially confounding factors. Therefore, multiple covariates assessed at T1 were included in the study design, based on work cited in the following sentences showing them to be related to one or more of the STR and problem behavior constructs central to this inquiry. For children's characteristics, we selected gender (Birch & Ladd, 1997) and language ability using a Norwegian adaptation of the Peabody Picture Vocabulary Test–III (10 items; Dunn & Dunn, 1997; O'Connor et al., 2011), Cronbach's alpha = .98. We also examined temperament (Rudasill, Reio, Stipanovic, & Taylor, 2010), specifically, negative affectivity. This measure is based on the anger, discomfort, fear, sadness, and soothability (reversed) scales (62 items), as reported by parents using the Children's Behavior Questionnaire for children 3–7 years of age (Rothbart, Ahadi, Hershey, & Fisher, 2001), Cronbach's alpha = .88. On the family level, covariates included level of maternal education (McCartney et al., 2010) in five categories (lower secondary, upper secondary, vocational school, college, university) and parental depression, assessed using the Beck Depression Inventory–II (21 items; Beck, Steer, & Brown, 1996; McCartney et al., 2010), Cronbach's alpha = .89.

Statistics

We applied structural equation modeling in Mplus (Muthén & Muthén, 2008), employing maximum likelihood estimation with robust standard errors. Since the sample was stratified at screening, analyses were weighted proportionally to the inverse of the probability of selection of each participating child. This provided unbiased general population estimates. Missing data were handled with full information maximum likelihood estimation.

Analytically, we first assessed reciprocal relations between measured variables (teacher–child conflict and closeness, child problem behavior, using sum scores) and stability in problem behavior over time in an autoregressive cross-lagged model. All T1 measures and all T2 measures were adjusted for all T1 covariates. This implied a fully saturated model that, by necessity, fits the data perfectly. Next, we tested whether associations were moderated by preschool group size—by means of a multigroup analysis, with significance levels of differences in paths tested by constraining one path at a time. Comparison of the differences between models was based on the corrected chi-square difference test (Satorra & Bentler, 2001).

Results

Means, standard deviations, and bivariate associations (correlations) between analyzed variables are displayed in Table 1. Child–teacher conflict and closeness were negatively associated within and across time. Behavior problems were negatively associated with closeness and positively associated with conflict.

Examining Reciprocal Relations

In the regression analysis, we first examined the effects of preschool student–teacher conflict as well as closeness and children's behavior problems on the respective outcomes (conflict, closeness, and behavior problems) in first grade, controlling for possible confounding child and family factors. The resulting autoregressive cross-lagged model is displayed in Figure 1. Stability in conflict and closeness in teacher–child relationships was evident, albeit moderate, even though there was a change of all teachers between day care and school. Higher preschool conflict scores predicted higher levels of problem behavior in first grade, whereas higher score on closeness predicted lower levels of subsequent problem behavior. Examining how early problems related to later teacher–child relationships revealed that children manifesting more preschool-age problem behavior experienced more conflict with their first-grade teachers, whereas no association

Table 1
Descriptive Statistics and Correlations Between Variables

Variable name, theoretical min & max	Min	Max	*M*	*SD*	1	2	3	4	5	6
1. Closeness, T1, 11–55	19	54	39.15	4.49						
2. Closeness, T2, 11–55	21	50	38.33	4.90	0.22**					
3. Conflict, T1, 12–60	12	45	17.88	4.70	−0.14**	−0.11**				
4. Conflict, T2, 12–60	12	46	18.02	4.86	−0.07	−0.17**	0.29**			
5. Behavioral problems, T1, 0–50	0	35	7.12	6.17	−0.09**	−0.09*	0.22**	0.15**		
6. Behavioral problems, T2, 0–70	0	30	3.93	4.55	−0.17**	−0.19**	0.25**	0.32**	0.60**	
7. Group size	3	55	18.9	6.4	−0.03	0.01	0.00	0.01	0.02	0.02
8. Gender (1 = boys)	1	2	1.5	0.5	0.07*	0.12**	−0.06	−0.14**	−0.08*	−0.15**
9. Language ability, 0–119	23	119	91.39	21.90	0.00	−0.08*	0.00	0.04	0.05	0.05
10. Negative affectivity, 1–7	1.7	5.1	3.7	0.5	−0.01	−0.01	0.09*	0.01	0.50**	0.28**
11. Maternal education, 1–5	1	5	3.8	1.11	0.08*	0.06	−0.12**	−0.10*	−0.13**	−0.11**
12. Parental depression, 0–63	0	39	4.4	5.3	−0.08*	−0.07	0.02	−0.07*	0.27**	0.21**

Note. T1 = Time 1; T2 = Time 2.
*$p < .05$. **$p < .01$.

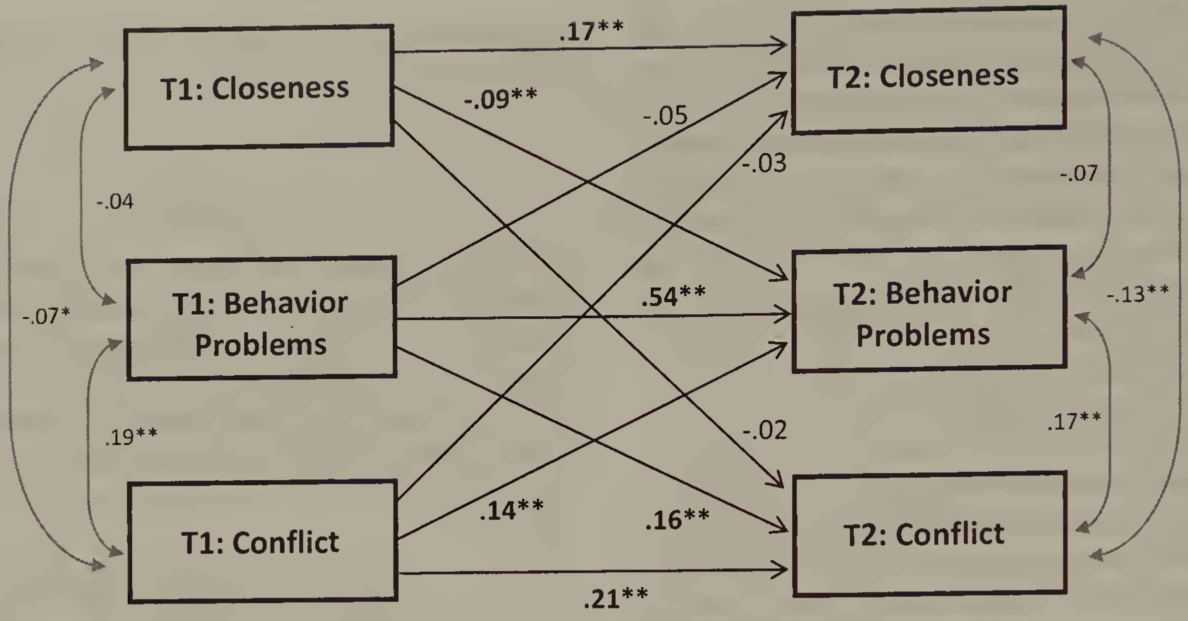


Figure 1. Main model with cross-lagged effects (standardized coefficients). Adjusted for covariates.
$^{*}p < .05$. $^{**}p < .01$.

between preschool-age problem behavior and later closeness with teachers emerged. Note that there was no direct effect of preschool group size on any of the outcomes, and potential effects of group size were also tested in the autoregressive cross-lagged model. R^2 coefficients were .37 ($p < .001$) for behavior problems, .10 ($p < .01$) for conflict, and .05 ($p = .01$) for closeness.

Moderation Effects of Group Size

Next we assessed whether the aforementioned bidirectional paths varied as a function of group size. Only some results proved consistent with predictions (Figure 2): In smaller groups (up to 15 children), child–teacher closeness at preschool predicted fewer behavior problems in first grade ($\beta = -.21$, $p < .001$), but as expected, no such effect of closeness on later behavior problems was evident in the case of children from larger groups ($\beta = -.05$, $p = .06$); critically, the coefficients for the two groups proved significantly different by corrected chi-square difference test, $\Delta\chi^2 = 4.23$ (1), $p = .04$. The effect of (greater) preschool child–teacher conflict on (more) problem behavior when children were in first grade was stronger in smaller groups ($\beta = .28$, $p = .01$) compared to larger groups ($\beta = .08$, $p = .05$), as expected, but the difference in coefficients across groups was not significant, $\Delta\chi^2 = 1.87$ (1), $p = .17$. Finally, the effect of preschool problems on child–teacher conflict in first grade was positive and larger in larger groups ($\beta = .18$, $p < .001$) than in smaller groups ($\beta = .11$, $p = .06$), but again the relevant coefficients did not prove to be significantly different by the corrected chi-square difference test, $\Delta\chi^2 = 0.20$ (1), $p = .65$.

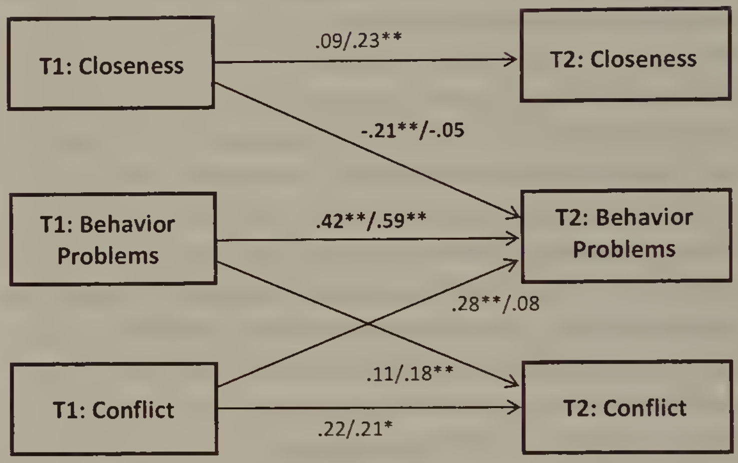


Figure 2. Moderation effects of group size (standardized coefficients by two categories of group size are presented: first by group size < 16, and second by group size 16 or greater). Significant interactions with group size are in bold.
$^{*}p < .05$. $^{**}p < .01$.

R^2 coefficients were .36 ($p < .001$) for behavior problems, .12 ($p = .11$) for conflict, and .02 ($p = .31$) for closeness in the smaller group. The respective measures in the larger group were .39 ($p < .001$) for behavior problems, .11 ($p = .01$) for conflict, and .08 ($p < .01$) for closeness.

Stability of Problem Behavior

As anticipated, problem behavior proved less stable over time among children from smaller groups than larger groups ($\beta = .42$, $p < .001$ and $\beta = .59$, $p < .001$, respectively), and this difference in stability coefficients proved to be statistically reliable by the corrected chi-square difference test, $\Delta\chi^2 = 5.47$ (1), $p = .02$. A simple slope analysis revealed that the effect of the moderated stability in problem behavior started to be significantly different by group size when children scored 9 or more on the externalizing behavior scale. This means that it was only children with increased behavioral problems who displayed significantly higher stability in behavior problems when attending large groups compared to small size groups ($F = 4.10$, $p = .04$). Because closeness and child behavior were correlated (Table 1), we could have been tapping into the same phenomena. Therefore, we included both moderations in the same model, and these were still significantly different from a model specifying no moderation, $\Delta\chi^2 = 10.72$ (2), $p < .001$.

Discussion

Consistent with prior work, the current child-care-related study in Norway revealed that a more conflicted relationship between preschool teacher and child predicted more behavior problems when children were of first-grade age, even after controlling for the stability of problem behavior, but that greater closeness to teachers forecast fewer behavior problems in school-aged children. In other words, when STRs were poorer prior to school entry, behavior problems increased more over time, with the reverse being the case when these relationships were more positive. Although the effect of closeness on later problem behavior was small, our study makes an important contribution to the existing and rather limited evidence concerning student–teacher closeness and externalizing behavior (see Zhang & Sun, 2011). In fact, our study is among the first to report significant effects of closeness as a predictor of externalizing behavior development.

Furthermore and consistent with the reciprocal-process hypothesis, more early problems predicted increased conflict with school teachers 2 years later. However, early behavior problems were not associated with future teacher–child closeness. This finding is consistent with prior research (Zhang & Sun, 2011) and can be attributed to the presumption that teacher–child conflict is to a larger degree driven by the teacher's perception of children's externalizing behavior, whereas closeness is more likely to be a function of the teacher's skills, sensitivity, and responsiveness (Silver et al., 2005).

The present study expands existing knowledge by suggesting that the reciprocal relations between children's behavior and relationship with their teacher persist through the transition from preschool to first grade. Our inquiry also showed that children's development related to two different social arenas—school and home environment—is interconnected. In particular, our findings underscore the importance of the preschool STR, which affects children's behavioral development beyond the context of an educational institution.

Additionally and consistent with our hypothesis regarding the moderating effect of preschool classroom context, specifically, group size, results indicated that experience in smaller groups strengthened the association linking greater preschool closeness with decreased problems 2 years later as compared to participation in larger groups; indeed, under the latter conditions, quality of teacher–child relationships in preschool failed to predict first-grade functioning. Although predictive coefficients relating the effect of conflict on later problem behavior were also stronger in smaller groups, consistent with expectations, this moderating effect of group size did not yield significantly different predictive coefficients across the two groups.

Recall that it was also predicted that participation in larger groups would generate greater stability in behavior problems across the transition to school. The data proved consistent with this hypothesis to a significant extent. However, with regard to the moderating effect of group size on the power of preschool problems to predict first-grade child–teacher relationships, the difference in coefficients across the two groups did not achieve statistical significance.

Although the results of the present study did not reveal any direct effect of group size on later children's functioning, contrary to some other prior research (see Dunn, 1993), our findings seem to highlight some benefits of small groups during the preschool period relative to larger groups. After all, supportive child–teacher relationships (i.e., greater closeness) in small groups forecast reduced growth of problems following the start of school and problem behavior in small groups displayed less stability. The latter result implies that children in small groups who manifest more problems than agemates were less likely to do so 2 years later relative to those with such problems in larger groups.

Why Might Small Groups Prove to Be a Beneficial Moderator?

We suggest that children might benefit from small groups due to the greater sense of security and warmth that a supportive teacher–child relationship affords in such contexts than in larger groups (Pianta, 1999; Pianta et al., 2003). Young children depend on adults to be told and shown appropriate behavior and also to have their behavior monitored and corrected when necessary (Jerome et al., 2009). It is possible that the beneficial impact of a close relationship with a teacher in smaller groups is enhanced by a larger degree of proximity, stability, predictability, and comparatively smaller degree of complexity in personal interactions. As a result, it might be easier for children in smaller than larger group settings to initiate interaction with their teacher and to make use of a teacher's support when needed. Similarly, teachers in small groups might be better positioned to provide individualized response to children's behavior. Such a conclusion might remain valid also under conditions of fixed teacher–child ratio, which arguably cannot substitute several aforementioned beneficial features of smaller groups compared to larger groups.

We thus contend that the beneficial effect of close relationships with teachers is enhanced in smaller groups due to the greater likelihood of positive one-to-one child–teacher interaction (Bourke, 1986), which might foster a greater sense of security. Such a perceived sense of emotional support might contribute to better self-regulation and internalization of classroom norms for appropriate behavior. Composition of groups, in terms of fewer children with problem behavior present in small groups compared to larger groups, might also influence children's behavior by providing more positive behavioral norms or expectations (Wright, Giammarino, & Parad, 1986). Additionally, small group membership (with fewer, but perhaps more frequent and intensive, relationships in the group) might strengthen the positive effect of a close STR by promoting children's feelings of predictability, support, emotional security, and confidence.

In larger groups, not only might positive and individual interactions with teacher be less frequent, but more negative and insensitive interchanges might take place more often. Increased complexity of interactions in larger groups might also make it more difficult for teachers to respond in a timely and sensitive manner to an individual child's behavior. In addition, peer influence, in terms of increased likelihood of observing and/or engaging in maladaptive behavior with both teachers and peers through observational learning, reinforcement learning, and peer-contagion processes (Dishion & Dodge, 2005) might account for why problem behavior proved more stable over time in the case of children who experienced larger rather than smaller preschool groups. Moreover, in such circumstances, problem behavior may come to be regarded as more normative and perhaps acceptable (Farmer, McAuliffe Lines, & Hamm, 2011). Unfortunately, this speculative analysis must remain just that, as proximal process measurements of quality of teacher–child and child–child interactions were beyond the measurement possibilities of this epidemiological study. This lacuna highlights at least one future direction for research while underscoring a fundamental limitation of the current inquiry.

The superiority of small group size can be explained by two main theoretical principles, as suggested by Finn, Pannozzo, and Achilles (2003) in their review of small classes at school (i.e., not in day care). According to the "visibility-of-the-individual principle," children in small groups cannot easily avoid being noticed and it is more difficult than in larger groups for teachers to ignore them. This makes it easier for teachers to monitor and respond in a timely manner to problematic behavior while being better positioned to notice—and reinforce—positive behavior. Additionally, the "sense-of-belonging principle" stipulates that small groups should foster greater cohesiveness and thus positive relationships among group members, including the provision and receipt of emotional support (Finn et al., 2003).

Importantly, our failure to detect any exacerbating effect of small group size on the relation between preschool child–teacher conflict and later behavior problems suggests that even though small groups apparently amplify effects of positive STRs, they do not strengthen effects of negative relationships, as we suspected might also be the case. This, of course, would seem to be good news. Possibly, management of comparatively fewer children in the group makes it easier for the teacher to respond in more appropriate ways to children's emotions and behavior, even when there is conflict.

In addition, it should be mentioned that teacher–child relationships contribute to the tenor of the relationship, but not to the frequency of interaction. Interpretations of the findings once again call attention to the need to measure carefully in future studies the dynamic social interaction processes that take place between teachers and children and among children in smaller and larger groups.

Limitations and Translational and Theoretical Implications

Recall that Norwegian law regulates many aspect of child care, including the adult–child ratio, but not group size. What the current research makes clear, then, is the potential benefit of regulating not just adult–child ratio because the apparent effects of group size cannot be attributed to differences in adult–teacher ratio. Unfortunately, we were not able to assess stability of day-care personnel, so cannot be certain that the role of group size discerned in this inquiry was not itself a function of staff absence or turnover. This is an important limitation of the research that needs to be highlighted because one recent Norwegian study found that in larger centers, which typically have more children per classroom, there is greater staff turnover and also staff absence due to illness (Vassenden et al., 2011). Such instability of personnel might undermine children's feeling of security, as well as reduce their comfort in interacting with or responding to teachers.

Another limitation is that children in child care all had at least another preschool teacher, but we collected information from only one teacher (who knew the child best). Thus, we cannot preclude the possibility that children might have different relationships with the other teacher. We might assume that additional relationships of the same quality in the child-care group might strengthen the observed relationship effect, while contrasting relationships might lessen it. In fact, we could speculate that the absence of the protective effect of closeness on problem behavior in large groups could be a function of other relationships, possibly of lower quality, in such groups. Finally, we have to acknowledge that various preschool as well as school factors, which we did not account for, such as years of teaching experience (O'Connor & McCartney, 2006), teacher's education (Howes, Whitebrook, & Phillips, 1992), teacher self-efficacy (O'Connor, 2010), classroom size, as well as child care and classroom environment (O'Connor, 2010; Pianta, 1999; Pianta et al., 2003; Sabol & Pianta, 2012), could have affected the reported relations. Future work should explore such possibilities.

Conclusion

The major contribution of this study involves extending consideration of group-size effects on the quality of child–teacher relationships and on children's development, two well-studied foci of developmental and educational interest, especially in the context of common teacher–child ratios across groups of different sizes. More specifically, in highlighting differences between smaller and larger groups in terms of the (apparent) influence of child–teacher relationships during preschool on the development and stability of behavior problems (in this observational/correlational study), this work highlights the need to think of group size in moderational, not just main-effect, terms, as has been the tradition in research on child care and classrooms. Moreover, due to the absence of regulations stipulating permitted group size, this work calls attention to the potential utility of regulating preschool group size in Norway—or eventually decreasing already regulated maximal group sizes in other countries.

References

Achenbach, T. M., & Rescorla, L. (2000). *Manual for the ASEBA preschool forms and profiles*. Burlington: University of Vermont.

Alvestad, T., Bergem, H., Eide, B., Johansson, J.-E., Os, E., Pálmadóttir, H., . . . Winger, N. (2014). Challenges and dilemmas expressed by teachers working in toddler groups in the Nordic countries. *Early Child Development and Care, 184*, 671–688. doi:10.1080/03004430.2013.807607

Baker, J. A. (2006). Contributions of teacher–child relationships to positive school adjustment during elementary school. *Journal of School Psychology, 44*, 211–229. doi:10.1016/j.jsp.2006.02.002

Bandura, A. (1971). *Psychological modeling: Conflicting theories*. Chicago, IL: Aldine Atherton.

Barth, J. M., Dunlap, S. T., Dane, H., Lochman, J. E., & Wells, K. C. (2004). Classroom environment influences on aggression, peer relations, and academic focus. *Journal of School Psychology, 42*, 115–133. doi:10.1016/j.jsp.2003.11.004

Beck, A., Steer, R., & Brown, G. (1996). *BDI–II, Beck Depression Inventory: Manual*. San Antonio, TX: Psychological Corporation.

Belsky, J., Vandell, D. L., Burchinal, M., Clarke-Stewart, K. A., McCartney, K., Owen, M. T., & the NICHD Early Child Care Research Network. (2007). Are there long-term effects of early child care? *Child Development, 78*, 681–701. doi:10.1111/j.1467-8624.2007.01021.x

Birch, S. H., & Ladd, G. W. (1997). The teacher–child relationship and children's early school adjustment. *Journal of School Psychology, 35*, 61–79. doi:10.1016/S0022-4405(96)00029-5

Birch, S. H., & Ladd, G. W. (1998). Children's interpersonal behaviors and the teacher–child relationship. *Developmental Psychology, 34*, 934–946. doi:10.1037/0012-1649.34.5.934

Blatchford, P. (2003). A systematic observational study of teachers' and pupils' behaviour in large and small classes. *Learning and Instruction, 13,* 569–595. doi:10.1016/S0959-4752(02)00043-9

Bourke, S. (1986). How smaller is better: Some relationships between class size, teaching practices, and student achievement. *American Educational Research Journal, 23,* 558–571. doi:10.3102/00028312013004558

Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *American Psychologist, 32,* 513–531. doi:10.1037/0003-D66X.32.7.513

Bronfenbrenner, U., & Morris, P. A. (1998). The ecology of developmental processes. In W. D. R. M. Lerner (Ed.), *Handbook of child psychology: Vol. 1. Theoretical models of human development* (5th ed, pp. 993–1028). Hoboken, NJ: Wiley.

Buyse, E., Verschueren, K., Doumen, S., Van Damme, J., & Maes, F. (2008). Classroom problem behavior and teacher–child relationships in kindergarten: The moderating role of classroom climate. *Journal of School Psychology, 46,* 367–391. doi:10.1016/j.jsp.2007.06.009

Clarke-Stewart, K. A. (1989). Infant day care: Maligned or malignant? *American Psychologist, 44,* 266–273. doi:10.1037/0003-D66X.44.2.266

Dishion, T., & Dodge, K. (2005). Peer contagion in interventions for children and adolescents: Moving towards an understanding of the ecology and dynamics of change. *Journal of Abnormal Child Psychology, 33,* 395–400. doi:10.1007/s10802-005-3579-z

Doumen, S., Verschueren, K., Buyse, E., Germeijs, V., Luyckx, K., & Soenens, B. (2008). Reciprocal relations between teacher–child conflict and aggressive behavior in kindergarten: A three-wave longitudinal study. *Journal of Clinical Child & Adolescent Psychology, 37,* 588–599. doi:10.1080/15374410802148079

Dunn, L. (1993). Ratio and group size in day care programs. *Child & Youth Care Forum, 22,* 193–226. doi:10.1007/BF00757426

Dunn, L. M., & Dunn, L. M. (1997). *Examiner's manual for the PPVT–III Peabody Picture Vocabulary Test.* Circle Pines, MN: American Guidance Service.

Fabes, R. A., Hanish, L. D., & Martin, C. L. (2003). Children at play: The role of peers in understanding the effects of child care. *Child Development, 74,* 1039–1043. doi:10.1111/1467-8624.00586

Farmer, T. W., McAuliffe Lines, M., & Hamm, J. V. (2011). Revealing the invisible hand: The role of teachers in children's peer experiences. *Journal of Applied Developmental Psychology, 32,* 247–256. doi:10.1016/j.appdev.2011.04.006

Finn, J. D., Pannozzo, G. M., & Achilles, C. M. (2003). The "why's" of class size: Student behavior in small classes. *Review of Educational Research, 73,* 321–368. doi:10.3102/00346543073003321

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry, 38,* 581–586. doi:10.1111/j.1469-7610.1997.tb01545.x

Hamre, B. K., & Pianta, R. C. (2001). Early teacher–child relationships and the trajectory of children's school outcomes through eighth grade. *Child Development, 72,* 625–638. doi:10.1111/1467-8624.00301

Howes, C., Phillipsen, L. C., & Peisner-Feinberg, E. (2000). The consistency of perceived teacher–child relationships between preschool and kindergarten. *Journal of School Psychology, 38,* 113–132. doi:10.1016/S0022-4406(99)00044-8

Howes, C., Whitebrook, M., & Phillips, D. (1992). Teacher characteristics and effective teaching in child care: Findings from the National Child Care Staffing Study. *Child & Youth Care Forum, 21,* 399–414. doi:10.1007/BF00757371

Hughes, J. N., & Cavell, T. A. (1999). Influence of the teacher-student relationship in childhood conduct problems: A prospective study. *Journal of Clinical Child Psychology, 28,* 173–184. doi:10.1207/s15374424jccp2802_5

Jerome, E. M., Hamre, B. K., & Pianta, R. C. (2009). Teacher–child relationships from kindergarten to sixth grade: Early childhood predictors of teacher-perceived conflict and closeness. *Social Development, 18,* 915–945. doi:10.1111/j.1467-9507.2008.00508.x

Kjelvik, J. (Ed.). (2012). *Barn og unges miljø og helse* [Children's and adolescents' milieu and health]. Oslo, Norway: Statistics Norway.

Laird, R. D., Jordan, K. Y., Dodge, K. A., Pettit, G. S., & Bates, J. E. (2001). Peer rejection in childhood, involvement with antisocial peers in early adolescence, and the development of externalizing behavior problems. *Development and Psychopathology, 13,* 337–354.

Legendre, A. (2003). Environmental features influencing toddlers' bioemotional reactions in day care centers. *Environment and Behavior, 35,* 523–549. doi:10.1177/0013916503035004005

Liew, J., Chen, Q., & Hughes, J. N. (2010). Child effortful control, teacher–student relationships, and achievement in academically at-risk children: Additive and interactive effects. *Early Childhood Research Quarterly, 25,* 51–64. doi:10.1016/j.ecresq.2009.07.005

Mantzicopoulos, P. (2005). Conflictual relationships between kindergarten children and their teachers: Associations with child and classroom context variables. *Journal of School Psychology, 43,* 425–442. doi:10.1016/j.jsp.2005.09.004

McCartney, K., Burchinal, M., Clarke-Stewart, A., Bub, K. L., Owen, M. T., & Belsky, J. (2010). Testing a series of causal propositions relating time in child care to children's externalizing behavior. *Developmental Psychology, 46,* 1–17. doi:10.1037/a0017886

Ministry of Education. (2012). *Til barnas beste* [For children's well-being]. Retrieved from http://www.regjeringen.no/nb/dep/kd/dok/nouer/2012/nou-2012-1/5.html?id=669130

Murray, C., Waas, G. A., & Murray, K. M. (2008). Child race and gender as moderators of the association between teacher–child relationships and school adjustment. *Psychology in the Schools, 45*, 562–578. doi: 10.1002/pits.20324

Muthén, L. K., & Muthén, B. O. (2008). *Mplus (version 5.1)*. Los Angeles, CA: Author.

National Institute of Child Health and Human Development Early Child Care Research Network. (2003). Does amount of time spent in child care predict socioemotional adjustment during the transition to kindergarten? *Child Development, 74*, 976–1005. doi:10.1111/1467-8624.00582

National Institute of Child Health and Human Development Early Child Care Research Network. (2006). Child-care effect sizes for the NICHD Study of Early Child Care and Youth Development. *American Psychologist, 61*, 99–116.

Norwegian Directorate for Education and Training. (2012). *Barn og ansatte i barnehager i 2012* [Children and employees in kindergartens in 2012]. Retrieved from http://www.udir.no/Barnehage/Statistikk-og-forskning/Statistikk/Barn-og-ansatte-i-barnehager-i-2012/Barn-i-ordinare-barnehager-familiebarnehager-og-apne-barnehager/

Nurmi, J.-E. (2012). Students' characteristics and teacher–child relationships in instruction: A meta-analysis. *Educational Research Review, 7*, 177–197. doi:10.1016/j.edurev.2012.03.001

O'Connor, E. (2010). Teacher–child relationships as dynamic systems. *Journal of School Psychology, 48*, 187–218. doi:10.1016/j.jsp.2010.01.001

O'Connor, E., Dearing, E., & Collins, B. A. (2011). Teacher–child relationship and behavior problem trajectories in elementary school. *American Educational Research Journal, 48*, 120–162. doi:10.3102/0002831210365008

O'Connor, E., & McCartney, K. (2006). Testing associations between young children's relationships with mothers and teachers. *Journal of Educational Psychology, 98*, 87–98. doi:10.1037/0022-0663.98.1.87

Pianta, R. C. (1999). *Enhancing relationships between children and teachers*. Washington, DC: American Psychological Association.

Pianta, R. C. (2001). *STRS Student–Teacher Relationship Scale. Professional manual*. Lutz, FL: Psychological Assessment Resources.

Pianta, R. C., Hamre, B., & Stuhlman, M. (2003). Relationships between teachers and children. In W. M. Reynolds & G. E. Miller (Eds.), *Handbook of psychology. Educational psychology* (Vol. 7, pp. 199–234). Hoboken, NJ: Wiley.

Pianta, R. C., & Stuhlman, M. W. (2004). Teacher–child relationships and children's success in the first years of school. *School Psychology Review, 33*, 444–458.

Rothbart, M. K., Ahadi, S. A., Hershey, K. L., & Fisher, P. (2001). Investigations of temperament at three to seven years: The Children's Behavior Questionnaire. *Child Development, 72*, 1394–1408. doi:10.1111/1467-8624.00355

Rudasill, K. M., Reio, T. G., Jr. Stipanovic, N., & Taylor, J. E. (2010). A longitudinal study of student–teacher relationship quality, difficult temperament, and risky behavior from childhood to early adolescence. *Journal of School Psychology, 48*, 389–412. doi:10.1016/j.jsp.2010.05.001

Sabol, T. J., & Pianta, R. C. (2012). Recent trends in research on teacher–child relationships. *Attachment & Human Development, 14*, 213–231. doi:10.1080/14616734.2012.672262

Sameroff, A. (1975). Transactional models in early social relations. *Human Development, 18*(1–2), 65–79. doi:10.1159/000271476

Sameroff, A. J., & Mackenzie, M. J. (2003). Research strategies for capturing transactional models of development: The limits of the possible. *Development and Psychopathology, 15*, 613–640. doi:doi:10.1017/S0954579403000312

Satorra, A., & Bentler, P. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika, 66*, 507–514. doi:10.1007/bf02296192

Seland, M. (2011). *Livet i den fleksible barnehagen: Muligheter og utfordringer i en barnehage i endring* [Life in the flexible day care: Possibilities and challenges in an evolving day care]. Oslo, Norway: Universitetsforlaget.

Silver, R. B., Measelle, J. R., Armstrong, J. M., & Essex, M. J. (2005). Trajectories of classroom externalizing behavior: Contributions of child characteristics, family characteristics, and the teacher–child relationship during the school transition. *Journal of School Psychology, 43*, 39–60. doi:10.1016/j.jsp.2004.11.003

Statistics Norway. (2013a). *Kindergartens, 2012, final figures*. Retrieved from http://ssb.no/utdanning/statistikker/barnehager/aar-endelige/2013-06-17

Statistics Norway. (2013b). *Households' payments in kindergartens, price level survey, 2012/2013*. Retrieved from http://www.ssb.no/en/utdanning/statistikker/barnegenniv

Sutherland, K. S., & Oswald, D. P. (2005). The relationship between teacher and student behavior in classrooms for students with emotional and behavioral disorders: Transactional processes. *Journal of Child & Family Studies, 14*, 1–14. doi:10.1007/s10826-005-1106-z

Thomas, D. E., Bierman, K. L., Powers, C. J., & the Conduct Problems Prevention Research. (2011). The influence of classroom aggression and classroom climate on aggressive-disruptive behavior. *Child Development, 82*, 751–757. doi:10.1111/j.1467-8624.2011.01586.x

UNICEF. (2008). *The child care transition. A league table of early childhood education and care in economically advanced countries* (Innocenti Report Card 8). Florence, Italy: UNICEF Innocenti Research Centre.

Vassenden, A., Thygesen, J., Bayer, S. B., Alvestad, M., & Abrahamsen, G. (2011). *Barnehagenes organisering og strukturelle faktorers betydning for kvalitet* [Organization of child care and the importance of structural factors for quality]. Stavanger, Norway: International Research Institute of Stavanger.

Vermeer, H. J., & van IJzendoorn, M. H. (2006). Children's elevated cortisol levels at daycare: A review and meta-analysis. *Early Childhood Research Quarterly, 21,* 390–401. doi:10.1016/j.ecresq.2006.07.004

Wichstrøm, L., Berg-Nielsen, T. S., Angold, A., Egger, H. L., Solheim, E., & Sveen, T. H. (2012). Prevalence of psychiatric disorders in preschoolers. *Journal of Child Psychology and Psychiatry, 53,* 695–705. doi:10.1111/j.1469-7610.2011.02514.x

Wright, J. C., Giammarino, M., & Parad, H. W. (1986). Social status in small groups: Individual–group similarity and the social "misfit." *Journal of Personality and Social Psychology, 50,* 523–536. doi:10.1037/0022-3514.50.3.523

Zhang, X., & Sun, J. (2011). The reciprocal relations between teachers' perceptions of children's behavior problems and teacher–child relationships in the first preschool year. *Journal of Genetic Psychology, 172,* 176–198. doi:10.1037/0022-3514.50.3.523

Child Development, September/October 2015, Volume 86, Number 5, Pages 1571–1587

A Developmental Analysis of Caregiving Modalities Across Infancy in 38 Low- and Middle-Income Countries

Marc H. Bornstein and Diane L. Putnick
Eunice Kennedy Shriver National Institute of Child Health and Human Development

Jennifer E. Lansford
Duke University

Kirby Deater-Deckard
Virginia Polytechnic Institute and State University

Robert H. Bradley
Arizona State University

Caregiving is requisite to wholesome child development from the beginning of life. A cross-sectional microgenetic analysis of six caregiving practices across the child's 1st year (0–12 months) in 42,539 families from nationally representative samples in 38 low- and middle-income countries is reported. Rates of caregiving varied tremendously within and across countries. However, caregiving practices followed one of two developmental trajectories: (a) greater proportions of caregivers read, told stories, and named, counted, and drew with each additional month of infant age, and (b) proportions of caregivers who played, sang songs, and took their infants outside increased each month from birth but reached an asymptote at 4–5 months. Rates and growth functions of caregiving have implications for infant care and development.

Which caregiving practices do infants' principal caregivers engage in? How do those practices change as infants develop? How are those practices distributed in low- and middle-income countries (LMIC) around the world? Are those practices the same or different for infant girls and boys? This study attempts to contribute answers to these four significant developmental questions for populations on which very little contemporary research exists.

Infant Caregiving Practices

To survive infants must have their biological requirements for nourishment and protection met, but to thrive infants also need their psychosocial essentials for mental growth and emotional well-being addressed. In every society, caregivers act in ways to meet the biological needs of infants as well as infants' other psychosocial requisites. Beyond nurturance and safety, therefore, active and enriching caregiving responsibilities fundamentally invest in young children's education and socialization. Just as infants must connect with others through socioemotional interactions, they must also learn about the world through didactic interactions.

To address our first question, this study focused on six specific infant caregiving practices that represent these two broad modalities—reading; telling stories; naming, counting, and drawing; playing; singing songs; and taking infants outside the home. Together, these kinds of caregiving practices stimulate infants to engage and understand the environment (e.g., by naming) and to participate in and model social relationships with others (e.g., by playing). Socioemotional caregiving focuses on the dyad and includes physical and verbal strategies caregivers use to engage infants interpersonally (Emde, 1992; Stern, 1985). Didactic caregiving turns outward from the dyad and consists of physical and verbal strategies caregivers use to stimulate infant attention, demonstrate, and the like. Many investigators have operationally distinguished and contrasted these modalities of interaction (Bornstein, Cote, Haynes, Suwalsky, & Bakeman, 2012; Reddy, 2008; Stern, 1985). Each domain of caregiving is developmentally significant in the sense that together they help lay foun-

---

We thank A. Hancioglu, Y. Ota, O. Petrovic, and N. Ulkuer. This research was supported by the Intramural Research Program of the NIH, NICHD.

Correspondence concerning this article should be addressed to Marc H. Bornstein, Child and Family Research, *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, Suite 8030 6705 Rockledge Drive, Bethesda, MD 20892-7971. Electronic mail may be sent to Marc_H_Bornstein@nih.gov.

Published 2015. This article has been contributed to by US Government employees and their work is in the public domain in the USA.
All rights reserved. 0009-3920/2015/8605-0018
DOI: 10.1111/cdev.12402

dations for children's future emotional, social, communicative, and cognitive competencies (e.g., Bornstein, 2015; Smith, Adamson, & Bakeman, 1988). For example, social sensitivity and responsiveness promote children's interpersonal competencies (e.g., De Wolff & van IJzendoorn, 1997), and educational interactions relate to children's developing instrumental competencies (e.g., Olson, Bates, & Bayles, 1984).

In their bioecological theory, Bronfenbrenner and Morris (2006) postulated that human development is the joint product of process, person, context, and time. In this formulation, processes refer to dynamic interactions that the developing person experiences, and "proximal processes"—face-to-face interactions with others and engagement in activities with materials in the microsystem—are hypothesized as particularly significant to children's development. The caregiving practices included in these two basic caregiving modalities instantiate proximal processes aimed at supporting development in children. Moreover, socioemotional and educational caregiving practices are likely universal in the sense that they characterize caregiver–infant interactions in a wide variety of cultures. Parenting infants is multidimensional, modular, and specific (Bornstein, 2006, 2015). Although particular caregiving practices (both within and across dimensions) may serve similar needs for children, they are largely independent of one another; thus, there is a need to identify and focus empirically on individual caregiving practices. To ensure a comprehensive understanding of the links between caregiving practices and infant development, therefore, multiple caregiving practices must be examined individually across age.

Infancy as a Time of Change

Infancy is a period of life widely recognized for the dramatic changes that take place in biology, sensation, perception, cognition, language, emotions, and temperament. Although the general pace and thoroughgoingness of development during the 1st year are noteworthy (Bornstein, Arterberry, & Lamb, 2014), some changes are incremental and others saltatory. How do the two modalities of caregiving change across the 1st year? There are several possible answers to our second question.

It could be that most caregiving practices bear no fixed relation to infant age. On this view, similar percentages of caregivers would utilize each of the caregiving practices throughout infancy with no systematic relation between caregiving and infant age. Alternatively, some caregiving practices might bear a linear relation to infant age. Some theorists view developmental change as gradual. That is, infants grow steadily in physiological regulation, cognitive awareness, and social appreciation as the 1st year progresses and reciprocally more and more caregivers will engage in certain caregiving practices as infants age across the year whereas fewer and fewer caregivers will engage in other caregiving practices as infants age. A third possibility is that caregiving bears a nonlinear relation to infant age. Still other theorists view developmental change in the infant as step- or stage-like. That is, the neonate is a rather different organism from the infant, as infants emerge on individual schedules from the state of *fetus ex utero* (Emde, Gaensbauer, & Harmon, 1976) and later achieve successive developmental milestones (e.g., starting to smile, reach, crawl, talk, and walk) around specific ages and over relatively short intervals. Many spectators of infant development have posited that sequences of qualitatively distinct "stages," "periods," "biobehavioral shifts," "levels," "phases," "fundamental psychological transitions," and the like characterize development in the 1st year (Bowlby, 1969; de Rosnay & Murray, 2012; Emde et al., 1976; Erikson, 1950; Freud, 1916–1917/1966; McCall, 1979; Piaget, 1952; Sander, 1962; Sroufe, 1997; Trevarthen & Hubley, 1978; Vygotsky, 1934/1962). The implication of the third view is that development of caregiving would proceed in steps or stages.

On the argument that different caregiving practices are more or less independent, it seems unlikely that all caregiving practices we studied here would fit a single growth function; rather, specific caregiving practices likely change in specific ways as infants develop. Guided by the developmental literature, we posited that infants' interest in the environment and learning about the physical world (through educational interactions with caregivers) would grow continually. As infants age, and intentionality and flexibility in behavioral organization emerge and consolidate, infants are increasingly active in exploring their environment. That is, the scope of babies' apperception continually broadens as they look to their surrounds and reach out and grasp, for example. Although infant socioemotional development proceeds incrementally until about midway in the 1st year, at that time reciprocal exchange (Sander, 1962), development from primary to secondary intersubjectivity (Trevarthen & Hubley, 1978), discriminating sociability (Sroufe, 1996), and relational competencies (de Rosnay & Murray, 2012) become manifest and stabilize. Infants' interest in face-to-face play peaks between

3 and 5 months. From 6 months onward they become more and more interested and fascinated by external stimuli as they are increasingly able to sustain joint object attention, explore the environment, and manipulate objects, resulting in changed play preferences toward more object play instead of face-to-face play (Power, 2000). In turn, caregivers typically perceive shifts in infants' abilities and in infants' responsiveness to their caregiving. These perceived shifts in infant capability and behavior engender corresponding shifts in how caregivers engage infants. Thus, around the middle of the 1st year infants settle into a mutual focus of interest connecting with others and engaging the social world. In shorter words, some caregiving growth functions would be linear and others nonlinear.

Measuring Change in Infancy

How can we best trace developmental changes in infant caregiving? Trial and error, connectionist models, and symbolic approaches to accomplishing everyday tasks assert that discovering and adopting new strategies occurs slowly. By contrast, some forms of learning, insight, and strategy acquisition follow more rapid change. Rates of uptake (the amount of time or experience before the frequency of a new strategy reaches an asymptotic level) vary depending on characteristics of the task, participants, and situations (Siegler, 2006).

Observations of emerging strategies are normally too widely spaced to yield detailed information about developmental processes. Assessing samples at wide intervals can make change appear more abrupt than sampling more frequently because the wider the sampling interval, the more stage-like change appears to be (Robinson, Adolph, & Young, 2004). However, microgenetic analyses yield granular as well as rough depictions of paths of change and therefore unique information about development because microgenetic methods share three main features: Observations span periods of rapid change (here infancy), the density of observations is high (here monthly across the 1st year), and observations are analyzed intensively (here through graphical presentation) with the goal of inferring underlying processes (Siegler, 2006). Densely sampling development during periods of rapid change provides the temporal resolution necessary to expose processes. In turn, more precise analyses of strategy discovery and uptake, of generalizations across strategies, and of precursors to strategy discovery and uptake require identifying when a strategy first emerges, and their emergence can be specified reasonably precisely only through microgenetic assessments.

In general, developmental studies of fine-grain data suffer a trade-off between numbers of observations and numbers of participants: The greater the number of observations, the smaller the number of participants, and vice versa. We overcame both shortcomings with large numbers of participants (> 42,000) studied in closely spaced (monthly) observations by employing a cross-sectional variant of the microgenetic design.

Infant Caregiving in Context

Many factors influence infant caregiving, infant development, and caregiver–infant relationships, but few studies to date have examined variation in infant caregiving in LMIC, our third question. As it happens, most of the world's very young children live in LMIC—an estimated 560 million children under 5 years of age (Grantham-McGregor et al., 2007)—but precious little is known about the caregiving they receive or how it matters (Tomlinson, Bornstein, Marlow, & Swartz, 2014; Walker et al., 2007). The contemporary developmental literature is demonstratively skewed to populations living in technologically advanced high-income countries (HIC). Moreover, the northern European and North American societies typically included in developmental science research tend to be highly similar: In them, families normally adhere to the same basic nuclear organization, and parents play the same fundamental roles and share many of the same primary goals for their children. Studies that employ a wider contextual lens therefore promise more penetrating insights into infant caregiving practices and how they are implicated in children's development. Such lessons illuminate how expansive or circumscribed are presumed universals. To advance our general understanding of caregiving and infant development, therefore, both must be studied in different contexts. This study uses data from the Multiple Indicator Cluster Survey (MICS3; United Nations Children's Fund [UNICEF], 2006a) a nationally representative and internationally comparable survey of households in 38 LMIC. The samples in this study differ in terms of national backgrounds, which form a powerful basis for conceiving of different caregiving norms and for implementing different practices in rearing infants, and the country samples we recruited are representative of their nation state.

We studied caregiving practice–infant age associations in different LMIC for theoretical as well as for practical reasons. Psychological universals are processes that are likely shared across peoples, cultures, and nations. Identifying and understanding psychological universals is theoretically significant in developmental and social science. "The discovery of genuine psychological universals entails the generalization of psychological findings across disparate populations having different ecologies, languages, belief systems, and social practices" (Norenzayan & Heine, 2005, p. 763). Here, we studied caregiving in 38 different LMIC, but we also aggregated caregiving across countries to identify putative universal patterns of infant caregiving.

Practically, this study centers on infant experiences in the 1st year of life. Caregiving practices in infancy are often a neglected area of study in LMIC, where the main areas of interest in the 1st years of life normally focus on survival, nutrition, growth, disease, and motor development rather than infants' mental and socioemotional experiences. However, research indicates that many caregiving effects established in the first years carry forward in long-term development (Bornstein, 2014a)—a condition that may even be more applicable to children in LMIC given that the resources available in high-income, technologically advanced societies may create more opportunities to compensate for early deficiencies in caregiving. Comparative study also helps to identify caregiving practices that are plastic to contextual conditions. Such identification in turn aids in targeting interventions. Finally, cross-national study identifies caregiver involvement in specific practices that inform as to how caregivers in different places view and treat infants. If caregivers do not think infants are capable of play or benefit from being read to, infant caregivers may be less inclined to engage in play with infants or read to them (Göncü & Gaskins, 2007) and thereby deprive infants of growth-promoting experiences.

As indicated earlier, bioecological theory characterizes development as a joint function of process, person, context, and time (Bronfenbrenner & Morris, 2006). Processes (already discussed) refer to dynamic interactions that the developing person experiences. The developing person includes characteristics and qualities, such as age, that interact with characteristics of the context to influence development. The contexts of child development are conceived in terms of a hierarchically organized and interlinked set of nested systems, with lower level, more proximal contexts (caregiver–infant interactions) nested within higher level, more distal contexts (nation). The LMIC reported here vary widely in terms of history and ideology, beliefs and values, social and economic situations, as well as other sociodemographic factors thought to influence caregiving. However, all these nations constitute LMIC (UNICEF, 2006b) defined with reference to the World Bank's (2014) system of classification of economies based on gross national incomes per capita, quality of life (life expectancy, literacy rates), and economic diversification (labor force, consumption).

Gender

Our fourth question asked whether patterns of change in caregiving across the 1st year in LMIC are similar or different for infant girls and boys. It is important to understand possible gender differences from early in the life course because even small differences in patterns of treatment can cumulate over time, and so fashion variation in trajectories of children's development (Abelson, 1985; Prentice & Miller, 1992). Indeed, developments in brain science make clear that even seemingly innocuous differences in early experience can have impacts on brain structure and function (Kolb & Gibb, 2011). Parents may rear their infant girls and boys similarly or differently. Generally, child gender is believed to broadly organize different descriptions, impressions, and expectations of children from the start of life (Bornstein, 2013). Classic "Baby X" studies (where the gender of the infant is not known to study participants) suggest that parents even conceive of and behave toward infants differently depending on whether they think they are interacting with a girl or a boy. Lytton and Romney (1991) examined differential parental socialization of girls and boys through meta-analysis of the psychological and developmental literatures. They included 172 studies conducted in North America and non–North American Western societies. A clear effect to emerge from their survey was systematic parent encouragement of gender-typed activities in children. In consequence, it could be that caregivers engage more with infant girls or boys, preferentially engage in different caregiving practices with infant girls and boys, or begin to engage infant girls and boys at different ages in different ways. Each possible pattern would produce gender-distinctive developmental trajectories of infant caregiving.

It is widely believed that the social contexts of many LMIC favor men over women (Morrison & Jütting, 2005; United Nations Development Programme [UNDP], 2011) and that gender differences and differential treatment of children dominate in most parts of the developing world. Given limited resources, parents must decide how to allocate those resources to children and what roles to assign children within the family. In some LMIC, parents may choose girls or boys because one or the other is perceived to have a greater need. For example, if boys are more likely to attend school than girls, parents may be more likely to read to sons than daughters or to spend time naming and counting with them. However, the lives of girls and boys, and women and men may vary substantially across LMIC; therefore, we may not assume that girls and boys in LMIC are treated systematically differently. Some data from health, caregiving, and discipline suggest that young girls and boys in LMIC are treated more similarly than heretofore expected (Bornstein, Putnick, Lansford, Deater-Deckard, & Bradley, in press).

The historically strong and continuing focus on gender differences tends to mitigate or even entirely overlook gender similarities and the large variances of many characteristics within each gender. In effect, the substantial overlap in distribution for males and females may be discounted in situations where small mean differences are observed. The gender differences perspective can exaggerate, reinforce, or even leverage gender differences. In contrast to the long-standing focus on gender differences, the "gender similarities" hypothesis states that females and males are actually similar on many biological and psychological indices (Hyde, 2005, 2014). Wide variation in caregiving practices within gender would lead to an expectation that caregiving of infants of the two genders is similar or that any gender differences are modest in size. Some early evidence for this hypothesis can be found in Maccoby and Jacklin's (1974) narrative review of parents' gender-differentiated socialization practices. Even granting that parents reinforce certain gender-typed play activities and toy choices, Maccoby and Jacklin (1974, p. 342) concluded that "the reinforcement contingencies for the two sexes appear to be remarkably similar." Much stronger evidence for a hypothesis of gender similarities emerged from a review of 46 meta-analyses of the psychological and developmental research literatures. All told, 124 effect sizes for gender comparisons across a range of psychological domains, including cognitive, communicative, social and personality, well-being, motor, and a heterogeneous variety of other developmental categories, yielded 78% of reported gender difference effect sizes that were small or close to 0 (Hyde, 2014).

There is too little programmatic research on how, at the family level, young girls and boys are reared, and what situations obtain across countries or only in select countries. At the outermost circle of developmental influences in the ecological perspective are overarching macrosystem patterns of beliefs, values, customs, and living conditions—culture, religion, the socioeconomic organization of society, and (most germane here) national indicators of gender equality. The overarching macrosystems context exerts profound effects on whether and how development is gendered, what is considered adaptive for each, which tasks girls and boys are prescribed and proscribed, and what roles they likely adopt as mature women and men. The macrosystem is not separate from children's more immediate environments; rather, it permeates and colors microsystems. Understanding the meaning and effect of proximal microsystem caregiver practices on the child often requires setting them within the broader macrosystem in which they occur (Bornstein, 1995).

In consequence, population-based multinational data from LMIC are also indispensable to identify the status, treatment, and condition of girls and boys in the majority world as well as how gender intersects with different domains of development more generally. Our understanding of child gender is severely limited by the existing body of research, and it is unclear whether gender lessons learned from the minority developed world apply to the majority developing world.

This Study

In overview, this report examines microgenetically the developmental distributions of six prominent caregiving practices across the 1st year of life toward girls and boys in more than 42,500 families in 38 LMIC. This work was guided by four main questions concerning those practices: (a) prevalence, (b) development, (c) context, and (d) gender. Answers to these questions will inform expected rates, developmental courses, and differentiation of infant caregiving practices in LMIC around the world.

Method

Participants

This study evaluates caregiving of singleton infants under 1 year of age in 42,539 families in 38 LMIC (see Table 1) using the MICS3. Families with

Table 1
Descriptive Statistics and Correlations With Infant Age by Country for Reading, Telling Stories, and Naming, Counting, and Drawing

		Reading books			Telling stories			Naming, counting, drawing		
Country	*n*	*M*	*SD*	r_{pb}	*M*	*SD*	r_{pb}	*M*	*SD*	r_{pb}
High HDI	3,698	.24	.43	.21***	.38	.49	.21***	.47	.50	.15***
Albania	177	.19	.40	.08	.19	.39	.07	.14	.35	.08
Belarus	506	.37	.48	.43***	.34	.47	.26***	.24	.43	.30***
Bosnia and Herzegovina	513	.18	.38	.19***	.38	.49	.23***	.72	.45	.14***
Kazakhstan	892	.14	.34	.19***	.22	.41	.24***	.46	.50	.10**
Macedonia	528	.30	.46	.10*	.48	.50	.15***	.86	.35	.03
Montenegro	170	.23	.42	.18*	.60	.49	.30***	.26	.44	.27***
Serbia	697	.19	.39	.24***	.50	.50	.25***	.20	.40	.24***
Trinidad and Tobago	215	.54	.50	.27***	.61	.49	.22***	.74	.44	.18**
Medium HDI	21,508	.09	.29	.11***	.19	.39	.17***	.28	.45	.11***
Bangladesh	5,832	.04	.19	.13***	.27	.44	.18***	.13	.33	.12***
Belize	159	.45	.50	.18*	.44	.50	.15	.62	.49	.16*
Cameroon	1,394	.01	.08	.05	.08	.26	.12***	.26	.44	.02
Djibouti	385	.06	.25	.06	.08	.28	.05	.09	.28	.03
Georgia	400	.29	.45	.21***	.40	.49	.26***	.64	.48	.22***
Ghana	698	.02	.15	.06	.06	.23	.05	.19	.40	.02
Guyana	474	.32	.47	.19***	.40	.49	.22***	.51	.50	.17***
Jamaica	268	.34	.48	.25***	.31	.46	.29***	.55	.50	.24***
Kyrgyzstan	588	.10	.30	.19***	.16	.36	.25***	.51	.50	.18***
Laos	840	.03	.16	.11**	.03	.17	.11***	.49	.50	.09**
Mauritania	1,844	.03	.17	.02	.09	.28	.11***	.32	.47	.09***
Mongolia	774	.07	.25	.24***	.03	.17	.13***	.07	.26	.19***
Suriname	419	.20	.40	.03	.31	.46	.09	.48	.50	.12*
Syria	2,015	.12	.32	.10***	.16	.37	.18***	.15	.36	.09***
Tajikistan	840	.05	.21	.05	.17	.37	.12***	.46	.50	.04
Thailand	1,842	.30	.46	.23***	.31	.46	.22***	.44	.50	.25***
Uzbekistan	1,034	.08	.27	.15***	.14	.35	.18***	.37	.48	.07*
Vanuatu	348	.40	.49	.12*	.54	.50	.11*	.80	.40	.12*
Vietnam	503	.04	.19	.14**	.11	.31	.16***	.27	.44	.39***
Yemen	851	.02	.13	−.02	.05	.21	.06	.10	.30	.23***
Low HDI	12,646	.04	.20	.06***	.14	.35	.09***	.40	.50	.05***
Burkina Faso	1,162	.00	.06	.03	.02	.14	−.02	.15	.35	.01
Central African Republic	2,074	.04	.20	.06**	.10	.30	.09***	.53	.50	.08***
Côte d'Ivoire	1,918	.01	.11	.03	.06	.23	.13***	.38	.48	.07***
Gambia	1,561	.03	.16	.10***	.05	.22	.14***	.28	.45	.09***
Guinea-Bissau	1,186	.03	.17	.03	.53	.50	.11***	.27	.44	.04
Nigeria	2,920	.10	.30	.08***	.20	.40	.14***	.54	.50	.05**
Sierra Leone	968	.04	.19	.12***	.08	.28	.13***	.53	.50	.03
Togo	857	.01	.10	.05	.07	.25	.09**	.24	.42	.09*
HDI N/A	4,687	.04	.20	.07***	.13	.33	.13***	.63	.48	.11***
Iraq	3,374	.04	.21	.08***	.09	.28	.15***	.52	.50	.15***
Somalia	1,313	.03	.16	.02	.24	.43	.14***	.91	.28	.09**
Total	42,539	.09	.28	.12***	.19	.39	.15***	.37	.48	.09***

Note. For all behaviors, the correlation with infant age was strongest for the high-HDI group followed by the medium-HDI group, and finally the low-HDI group (we did not compare relations for the HDI N/A group). HDI = Human Development Index.
$^{*}p \leq .05$. $^{**}p \leq .01$. $^{***}p \leq .001$.

multiple children under 1 year (e.g., twins and triplets) were excluded in order to have independent observations and because parents with multiple infants may behave differently than do parents with single infants (Bornstein & Ruddy, 1984; Lytton & Gallagher, 2002). Infant age was coded as the month and year of assessment minus the month and year of birth, resulting in 12 monthly categories ranging from 0 months (1st month of life) to 11 months (12th month of life). The sample included 3,100–3,700 children in each monthly age group, and girls and boys were equally distributed across each month of age, $\chi^2(11, N = 42{,}536) = 11.78$, $p = .381$. Of those contacted to participate in the study, 96.6% completed the interview (range = 88.1%–100.0%) across countries, 2.7% were not at home, 0.2% refused, and 0.5% did not participate for other reasons. Infants averaged 5.45 months of age ($SD = 3.39$, range = 0–11), and 49.3% were girls. MICS questions were answered by the infant's primary female caregiver, who was almost always the infant's biological mother (97.6%). Of the 2.4% of questionnaires that were completed by someone other than the infant's biological mother, 89.6% had no biological mother living in the household. The child's primary female caregiver averaged 27.11 years of age ($SD = 7.10$, range = 10–84), which varied across countries, $F(37, 42{,}485) = 49.65$, $p < .001$, $\eta^2 = .041$. Caregiver education also varied, $\chi^2(111, N = 42{,}498) = 23{,}805.15$, $p < .001$, with 32.1% of caregivers completing no schooling or only preschool, 30.1% completing primary/religious school, 32.6% completing secondary/vocational/tertiary school, and 5.3% completing higher education.

Procedures

The 1990 World Summit for Children adopted the World Declaration on the Survival, Protection, and Development of Children and its Plan of Action, where governments pledged to monitor progress toward achieving those goals. To aid in this effort, UNICEF developed the MICS, a nationally representative and internationally comparable household survey. The third round of the MICS (MICS3) on which these analyses are based was carried out between 2005 and 2010. Each country designed and selected a national probability sample and field implemented the MICS3 with minimum deviation from an overall standard design. A three-stage sampling frame was used: (a) primary sampling units were defined, if possible, as census enumeration areas, and they were selected with systematic probability proportionate to size (pps); (b) segments (clusters) were designated; and (c) households were selected within each segment that were to be interviewed in the survey. To simplify implementation, implicit stratification was followed. When this form of geographic stratification is used together with pps sampling, the sample proportionately distributes into each of a nation's administrative subdivisions as well as its urban and rural sectors. Participants were visited at home and interviewed by trained, local field workers.

Caregiving

The questions used here were part of the optional child development module of the MICS3 Under Five Questionnaire (for details, see Bornstein et al., 2012). This study included responses from female caregivers about what mothers, fathers, and other caregivers did with infants in the past 3 days. The six MICS3 items included whether mother, father, or other caregiver (a) read books to the infant; (b) told the infant stories; (c) named, counted, or drew with the infant; (d) played with the infant; (e) sang songs to the infant; or (f) took the infant outside the home, yard, compound, or enclosure. Items were coded as 0 = no, 1 = yes. We included all caregivers because nonparents (siblings, grandparents, community members) may take active roles in infant caregiving in LMIC (Bornstein, 2015; Leinaweaver, 2014; Smith & Drew, 2002). Items went through forward-translation of questionnaires (available in English, French, Spanish, Russian, and Arabic) into (major) local languages and then back-translation by a separate translator without referring to the original questionnaires. (More details about translation appear in the MICS3 manual, p. 3.8; http://www.childinfo.org/mics3_manual.html.)

Country-Level Development

The countries in this study were organized according to the Human Development Index (HDI; UNDP, 2008). The HDI has three major components: life expectancy (in years), education (composed of the adult literacy rate and the percentage of school-aged children enrolled in primary, secondary, and tertiary school), and gross domestic product (in purchasing power parity in U.S. dollars). The HDI offers a proxy for the level of support that is generally available to promote human development in a country. The HDI ranges from 0.00 to 1.00, and countries with an HDI of 0.90 to 1.00 are considered very high, 0.80 to 0.89 are considered high, 0.50 to 0.79 medium, and 0.00 to 0.49 low. None of the very high HDI countries meet criteria to be classified as low- or middle-income, so this sample did not include any very high HDI countries. Due to missing country-level data, the HDI was not available for two countries. Because MICS3 data were collected between 2005 and 2010, we used the 2008 HDI (UNDP, 2008), which is based on 2006 data.

Analytic Plan

First, basic data are described. Then, in accordance with principles of microgenetic analysis, data were visualized by graphing the proportions of caregivers who engaged in each caregiving practice against infant age in months. For dichotomous variables, such as strategy discovery or uptake, logistic regression can be used to determine hazard rate, that is, the probability of an event occurring. Next, for caregiving practices that had linear relations with infant age, point biserial correlation and logistic regression analyses including infant gender as a moderator were performed across all countries and for each country separately. For caregiving practices that appeared to have nonlinear relations with infant age, a family of logistic spline regression analyses (Marsh & Cormier, 2002) was performed overall and by country to determine if and where on the infant age continuum the logit function changed (i.e., akin to a change in slope or a "knot" in the regression line). Logistic spline regression (aka piecewise logistic regression) allows for different logit functions (pieces) in segments of the sample. The average knot ages across countries (excluding out-of-range estimates—those that were negative or were greater than 12 months, indicating that there was no knot within the age range we studied in the country) were then used in logistic regression analyses to determine whether the logit functions differed significantly when the distribution was segmented at the knot age. Infant gender was included as a potential moderator to determine if changes in logit functions differed for girls and boys. Last, based on the average knot locations across countries, point-biserial correlation analyses were performed on younger and older infants to clarify the piecewise logit functions.

Results

Prevalences of Caregiving Practices

Tables 1–4 show the average proportions (means and standard deviations) of caregivers who engaged in each kind of caregiving overall and by country (Question 1). Proportions of caregivers engaging in each varied across caregiving practices. For example, only 9% of caregivers had read to their infants in the past 3 days, but 70% of caregiv-

Table 2

Summary of Spline Logistic Regression for Playing by Country

Country	*M*	*SD*	Knot estimate	r_{pb} (0–5 month)	r_{pb} (6–11 month)
High HDI	.83	.37	4.000	**.35*****	**.08*****
Albania	.73	.44	5.000	**.46*****	**−.01**
Belarus	.94	.23	4.000	**.40*****	**.07**
Bosnia and Herzegovina	.93	.26	1.071	.32***	.08
Kazakhstan	.76	.43	8.000	**.39*****	**.09**
Macedonia	.82	.39	−1.333	.24***	.15*
Montenegro	.89	.32	11.059	.51***	.26*
Serbia	.76	.43	6.749	**.37*****	**.07**
Trinidad and Tobago	.97	.18	−4.658	−.09	.07
Medium HDI	.67	.50	4.482	**.25*****	**.05*****
Bangladesh	.66	.47	4.480	**.26*****	**.03**
Belize	.92	.26	11.490	.16	.18
Cameroon	.80	.40	4.018	**.33*****	**.06**
Djibouti	.11	.32	−8.223	.04	−.01
Georgia	.83	.38	3.799	**.38*****	**.10**
Ghana	.84	.36	1.277	**.30*****	**.06**
Guyana	.86	.34	2.543	**.29*****	**.02**
Jamaica	.94	.25	4.232	.27***	.14
Kyrgyzstan	.50	.50	7.869	**.27*****	**.05**
Laos	.53	.50	3.000	.14**	.03
Mauritania	.46	.50	4.544	**.34*****	**.06**
Mongolia	.53	.50	7.000	**.34*****	**.10**
Suriname	.89	.31	4.000	.18*	.08
Syria	.65	.48	6.457	**.25*****	**.03**
Tajikistan	.58	.49	8.743	**.25*****	**.11***
Thailand	.92	.27	3.586	**.29*****	**.08***
Uzbekistan	.55	.50	6.484	**.31*****	**.07**
Vanuatu	.86	.35	6.000	.22**	.00
Vietnam	.82	.39	4.000	.25***	.13*
Yemen	.61	.49	6.647	**.26*****	**.10***
Low HDI	.71	.45	4.108	**.20*****	**.04*****
Burkina Faso	.56	.50	4.384	**.29*****	**.01**
Central African Republic	.61	.49	6.405	**.18*****	**.00**
Côte d' Ivorie	.85	.35	4.735	**.22*****	**.06**
Gambia	.56	.50	3.310	**.29*****	**.13*****
Guinea-Bissau	.52	.50	1.651	**.18*****	**.09***
Nigeria	.81	.39	2.798	**.17*****	**.05**
Sierra Leone	.96	.20	1.000	.13**	.10*
Togo	.76	.43	4.000	**.23*****	**−.01**
HDI N/A	.68	.47	3.678	**.28*****	**.06*****
Iraq	.68	.47	3.405	**.28*****	**.08****
Somalia	.69	.46	4.323	**.27*****	**.03**
Total	.70	.46	4.916[a]	**.24*****	**.05*****

Note. Correlations that significantly changed at 5 months of age are bolded. HDI = Human Development Index.

[a] Average of individual country knots excluding out-of-range estimates.

$^{*}p \leq .05$. $^{**}p \leq .01$. $^{***}p \leq .001$.

Table 3
Summary of Spline Logistic Regression for Singing by Country

Country	*M*	*SD*	Knot estimate	r_{pb} (0–4 month)	r_{pb} (5–11 month)
High HDI	.81	.39	3.259	**.12*****	**.04**
Albania	.82	.39	−1.172	.04	.06
Belarus	.84	.36	17.083	.13	.14**
Bosnia and Herzegovina	.67	.47	5.001	.06	.06
Kazakhstan	.88	.32	4.394	.13*	.08
Macedonia	.75	.43	5.503	**.16***	**−.03**
Montenegro	.96	.20	2.361	.39**	.09
Serbia	.76	.42	3.675	**.21*****	**.00**
Trinidad and Tobago	.91	.29	−4.534	−.03	.05
Medium HDI	.54	.50	1.461	**.16*****	**.04*****
Bangladesh	.37	.48	1.554	**.18*****	**.07*****
Belize	.77	.42	1.000	.22	.19
Cameroon	.62	.49	7.000	.15***	.07
Djibouti	.12	.33	−1.419	.03	−.00
Georgia	.74	.44	1.529	.19*	.01
Ghana	.68	.46	2.000	.15**	.07
Guyana	.78	.42	4.261	**.23*****	**−.02**
Jamaica	.82	.38	3.147	.21*	.09
Kyrgyzstan	.66	.47	1.460	**.20*****	**−.00**
Laos	.10	.30	6.000	.08	−.01
Mauritania	.51	.50	6.302	**.29*****	**.09****
Mongolia	.50	.50	7.000	.07	.06
Suriname	.79	.41	6.551	**.18***	**−.00**
Syria	.64	.48	7.001	**.12*****	**.02**
Tajikistan	.66	.47	4.996	.04	.04
Thailand	.78	.41	4.807	**.17*****	**.01**
Uzbekistan	.78	.41	7.000	.10	.07
Vanuatu	.78	.41	4.695	**.19*****	**−.02**
Vietnam	.30	.46	8.000	.15*	.20***
Yemen	.39	.49	3.000	**.19*****	**.05**
Low HDI	.54	.50	3.623	**.14*****	**.03*****
Burkina Faso	.18	.38	5.564	**.09**	**.01**
Central African Republic	.66	.47	2.413	**.18*****	**.02**
Côte d' Ivorie	.55	.50	4.701	**.11****	**.01**
Gambia	.30	.46	4.360	**.15*****	**.08***
Guinea-Bissau	.53	.50	1.062	**.15*****	**.06**
Nigeria	.65	.48	3.374	**.22*****	**.05***
Sierra Leone	.77	.42	7.949	.13**	.06
Togo	.75	.43	2.000	.14**	.07
HDI N/A	.49	.50	4.304	**.15*****	**−.02**
Iraq	.42	.49	6.035	**.14*****	**.01**
Somalia	.68	.47	1.494	**.15*****	**.05**
Total	.56	.50	4.329[a]	**.15*****	**.03*****

Note. Correlations that significantly changed at 4 months of age are bolded. HDI = Human Development Index.
[a] Average of individual country knots excluding out-of-range estimates.
*$p \leq .05$. **$p \leq .01$. ***$p \leq .001$.

Table 4
Summary of Spline Logistic Regression for Taking Outside by Country

Country	*M*	*SD*	Knot estimate	r_{pb} (0–4 month)	r_{pb} (5–11 month)
High HDI	.83	.37	3.860	**.31*****	**.10*****
Albania	.70	.50	1.000	.35***	.20
Belarus	.89	.31	7.666	.20*	.06
Bosnia and Herzegovina	.94	.23	1.000	.27***	.11*
Kazakhstan	.83	.37	3.174	**.31*****	**.07**
Macedonia	.81	.39	3.446	.27***	.13*
Montenegro	.85	.36	12.009	.49***	.21*
Serbia	.74	.44	7.099	**.35*****	**.11***
Trinidad and Tobago	.92	.27	6.036	.13	.06
Medium HDI	.71	.46	3.036	**.28*****	**.05*****
Bangladesh	.76	.43	2.141	**.34*****	**.06*****
Belize	.86	.35	7.410	.26*	−.06
Cameroon	.70	.46	3.102	**.34*****	**.04**
Djibouti	.11	.31	21.440	.04	.07
Georgia	.94	.25	−2.099	**.33*****	**.04**
Ghana	.75	.43	1.701	**.29*****	**.03**
Guyana	.83	.38	3.341	**.41*****	**.01**
Jamaica	.88	.32	3.778	**.33*****	**−.02**
Kyrgyzstan	.71	.45	6.662	**.24*****	**.01**
Laos	.58	.49	3.000	**.21*****	**.04**
Mauritania	.42	.49	4.554	**.21*****	**.03**
Mongolia	.78	.42	7.000	.15**	.08
Suriname	.83	.38	3.928	**.26*****	**.01**
Syria	.68	.47	4.222	**.26*****	**.05**
Tajikistan	.73	.44	3.937	**.24*****	**.05**
Thailand	.90	.30	2.879	**.33*****	**.08****
Uzbekistan	.86	.35	1.676	**.35*****	**.15*****
Vanuatu	.90	.31	2.498	.23**	.02
Vietnam	.55	.50	3.188	**.37*****	**.17****
Yemen	.39	.49	8.863	**.28*****	**.12****
Low HDI	.69	.46	3.189	**.19*****	**.04*****
Burkina Faso	.72	.45	6.445	**.14****	**−.06**
Central African Republic	.64	.50	6.091	**.20*****	**−.04**
Côte d' Ivoire	.88	.33	1.350	**.26*****	**.09****
Gambia	.64	.48	6.394	.19***	.16***
Guinea-Bissau	.34	.47	5.000	.08	−.00
Nigeria	.69	.46	3.223	**.20*****	**.07****
Sierra Leone	.85	.35	.829	.22***	.13**
Togo	.70	.46	1.386	**.31*****	**.04**
HDI N/A	.38	.49	6.053	**.22*****	**.08*****
Iraq	.39	.49	3.215	**.26*****	**.08*****
Somalia	.37	.48	6.886	**.11****	**.05**
Total	.68	.47	4.118[a]	**.25*****	**.06*****

Note. Correlations that significantly changed at 4 months of age are bolded. HDI = Human Development Index.
[a] Average of individual country knots excluding out-of-range estimates.
*$p \leq .05$. **$p \leq .01$. ***$p \leq .001$.

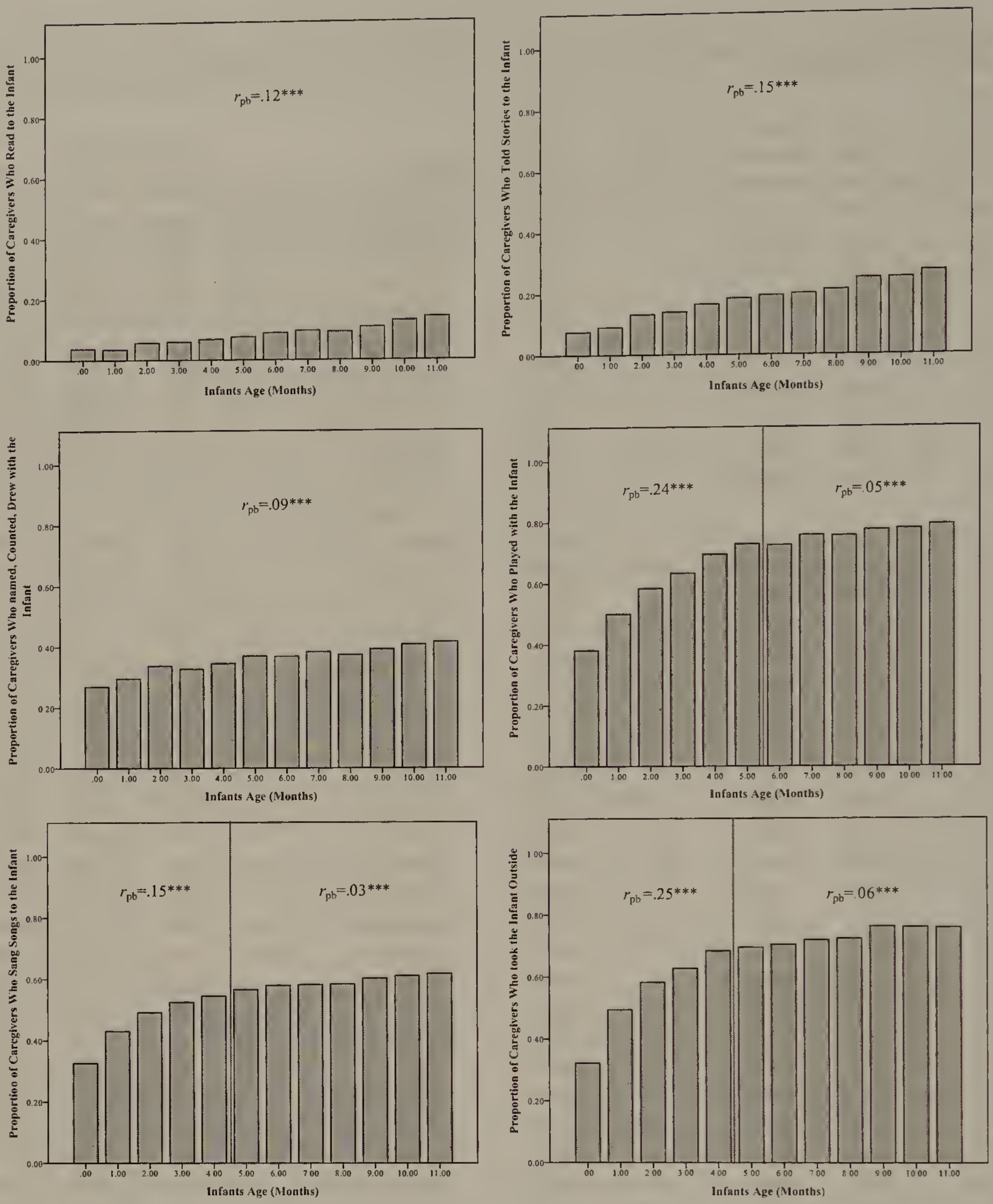


Figure 1. Proportions of caregivers who engaged in caregiving behaviors across the 1st year of life. Age in months is coded into categories such that, for example, 1 month of age includes children who are in their 2nd month of life. The break line indicates the knot age where the logit function changes.
***$p < .001$.

ers had played with their infants (see final rows in Tables 1–4). Despite large proportions of caregivers engaging in some practices, none of these caregiving practices displayed ceiling effects (Figure 1). The six kinds of caregiving also shared only 4%–27% of their variance, which supports exploring developmental functions in each caregiving practice separately.

Notably, 15% of infants reportedly received none of the six caregiving practices, and only 6% received all six caregiving practices in the past 3 days. Between-country variation was large. For example, virtually no caregivers (0%) read to infants in Burkina Faso whereas more than half (54%) of caregivers read to infants in Trinidad and Tobago. Wide ranges were also evident for telling stories (3%–61%); naming, counting, and drawing (7%–91%); playing (11%–97%); singing (10%–91%); and taking the infant outside (11%–94%), and proportions tended to be largest in high-HDI countries and smaller in medium- and low-HDI countries. Within-country variation was also substantial for most reported caregiving practices (see the standard deviations in Tables 1–4).

Prevalence Rates of Caregiving Practices Across Infancy, Among LMIC, and for Girls and Boys

Inspection of cross-sectional microgenetic graphs indicated that relations between infant age and the proportions of caregivers reading, telling stories, and naming, counting, and drawing were linear, but relations between infant age and the proportions of caregivers playing, singing songs, and taking their infant outside were nonlinear (Question 2; Figure 1). For all three of the logistic regressions of the linear caregiving practices (reading, telling stories, and naming, counting, and drawing), country interacted with infant age, Wald(37) = 84.41–230.60, *ps* < .001, indicating that proportions of caregivers who engaged in each practice across infant age were variable across countries (Question 3). Point biserial correlations of infant age with the proportions of caregivers reading, telling stories, and naming, counting, and drawing by country are presented in Table 1. The proportions of caregivers reading, telling stories, and naming, counting, and drawing were significantly positively related to infant age in 25 (66%), 31 (82%), and 29 (76%) of 38 countries, respectively. Correlations with infant age were strongest in high-HDI countries, followed by medium-HDI countries, and finally low-HDI countries (*zs* = 2.29–7.27, *ps* < .05). However, all effects were small to medium sized. Infant gender did not interact with age for any caregiving practices, Wald(1) = .01–1.85, *ps* = .173–.909, ORs = 1.00–1.01, indicating that the logit functions were similar across age for girls and boys (Question 4).

To determine the age at which the logit functions changed for the proportions of caregivers playing, singing songs, and taking the infant outside, we conducted a series of spline logistic regression analyses using 6 months as the start value for a single knot for each country. As shown in Tables 2–4, the estimated knot ages varied by country. The location of a single knot was estimated at 4–5 months (e.g., the average knots across countries = 4.12–4.92 months). In preliminary logistic regression analyses, a significant interaction emerged between the change in knot age and country, Wald(37) = 71.20–260.38, *ps* < .001, indicating that the change in logit functions was not consistent at the knot age across countries (Question 3). Consequently, we tested whether the logit function changed at the average knot age separately in each country. Using a knot of 5 months for playing, the logit function significantly flattened for 26 (68%) of 38 countries (bolded values in Table 2). Using a knot of 4 months for both singing songs and taking their infant outside, the logit functions significantly flattened for 19 (50%) and 25 (66%) of 38 countries (bolded values in Tables 3 and 4) respectively. When we aggregated the countries by HDI group, the logit functions for all three caregiving practices significantly flattened for high-, medium-, and low-HDI groups, indicating that the nonlinear developmental function was evident at all levels of country human development (Tables 2–4). Logit functions did not change at the knot age on any item for Belize, Bosnia and Herzegovina, Djibouti, Montenegro, Sierra Leone, and Trinidad and Tobago (16% of countries). Again, no significant interactions with infant gender were found, Wald(1) = .11–.72, *ps* = .398–.740, ORs = 1.00–1.01, indicating that the nonlinear functions were similar for girls and boys (Question 4).

Discussion

Healthy human infants cannot grow up without competent and engaging caregivers (Winnicott, 1964). Caregivers are responsible for infants' developmentally important experiences because infants themselves have limited capacities. Here, we report cross-sectional microgenetic analyses of six prominent infant caregiving practices in over 42,500 families in 38 LMIC (populations on which there is little research). Current knowledge and policy pertaining to infant care and development is framed by a body of research largely conducted on populations in HIC where a relatively small minority of children actually reside (Bornstein et al., 2014; Tomlinson et al., 2014). More needs to be learned about infant development and experience in the developing world of LMIC—where most families live. Moreover, studying microgenetic associations between caregiving practices and infant age in a range of diverse countries helps to identify socialization mechanisms and trajectories of their development that may be universal and submits the generalizability of such findings to test.

Infant Caregiving in LMIC

Caregiving practices with infants are not random behaviors that do not imply a particular goal, but more resemble nonobligatory goal-directed strategies. Many factors influence infant caregiving, infant development, and caregiver–infant relationships, but few studies to date have examined international variation in infant caregiving in LMIC (Walker et al., 2007). In answer to our first question about which caregiving practices infants' principal caregivers engage in, overall more caregivers of infants in every country played, sang songs, and took their infants outside than read, told stories, and named, counted, and drew with their infants, but there were tremendous ranges for these caregiving practices within and across countries. Taking an infant outside the home was the most prevalent, followed by playing, singing, naming, telling stories, and finally reading books. In the 1st year of life, infants' cognitive abilities are relatively limited, and didactic kinds of caregiving in terms of three MICS items (reading, storytelling, and naming, counting, or drawing) remained at a basal level with fewer than 40% of caregivers reporting each activity at any age and fewer than 10% of caregivers reportedly reading and telling stories at the youngest ages.

Rates for certain caregiving practices in LMIC may be low for a variety of reasons. Challenging in optimal circumstances, caregiving is even more difficult when family, community, and national resources are inadequate (Edin & Lein, 1997; Magnuson & Duncan, 2002). Numerous noxious links exist between economic privations on the one hand and parenting (and child development and well-being) on the other. For example, in the perspective of the "family stress model" (Conger & Donnellan, 2007), severe or persistent impoverishment stresses caregivers (e.g., stemming from day-to-day struggles to secure household resources and trying to cope with living in deteriorated or dangerous circumstances), which in turn undermine involvement with children and effective caregiving (see also Leventhal, Dupéré, & Shuey, 2015). These effects are likely compounded as a consequence of generally low parent literacy. Compared to middle-socioeconomic-status (SES) parents, low-SES parents even in HIC are less likely to converse, read, or provide appropriate play materials in the home to young children (Bradley, Corwyn, McAdoo, & Garcia Coll, 2001; Hoff, Laursen, & Tardif, 2002). Moreover, in LMIC young children are more likely to be sick and malnourished, which diminishes the likelihood that they will be recipients of less survival-essential kinds of caregiving. In situations where basic health concerns are deemed critical, it could also be that some parents delay focusing on nurturing psychosocial competencies in children pending assurance of the continued existence and health of their child. The interplay of these issues is strong in LMIC. Young Lives, a four-country study of child poverty, revealed that children living in low-resource contexts tend to exhibit impaired cognitive and socioemotional development (Dercon & Krishnan, 2009). Positive caregiving promotes infants' cognitive and socioemotional competencies. Consider book reading, a caregiving behavior that was relatively neglected across the LMIC we studied. Shared book reading stimulates infants' interests; exposes infants to sounds, print, and vocabulary; and provides opportunities for caregivers to create novel interactive learning environments with infants. The frequency with which adults read to children and the frequency of maternal labeling questions while reading books together (e.g., "What do you call this little animal?") correlate positively with children's vocabulary (Quiroz, Snow, & Zhao, 2010). The low proportions of children's principal caregivers who engaged in book reading in these LMIC contrast with findings in some HIC. Given that we compiled the caregiving of mothers, fathers, and other caregivers, the amount of children's exposure to reading in LMIC is quite low in comparison to the United States, for example, where 70% of 4- to 9-month-old infants are read to at least three times per week and only 9% are never read to (Kuo, Franke, Regalado, & Halfon, 2004).

Developmental Changes in Infant Caregiving

In answer to the second and third questions about how caregiving practices change as infants develop and how those practices are distributed in LMIC around the world, we found that the likelihood that parents would enact many of the caregiving practices examined grew incrementally across the 1st year of life; specifically, marginally more caregivers of infants engaged in didactic activities, such as reading, telling stories, and naming, counting, and drawing, with each additional month of infant development. This pattern was supported in 25–31 of 38 developing countries, with 89% (34 of 38) of countries showing a significant linear pattern for at least one of these kinds of didactic caregiving. Moreover, relations between caregiving and infant age were strongest in high-HDI countries followed by medium- and then low-HDI countries. One possible reason for the stronger relation in high-HDI countries may be that there was less restriction of range in high-HDI countries. For example, 38% of high-HDI caregivers told stories to infants (*SD* = 49, range = 19%–61%) compared to 19% of medium-HDI (*SD* = 39, range = 3%–54%) and 14% of low-HDI (*SD* = 35, range = 2%–53%) caregivers. Caregivers in high-HDI countries may also be more likely to begin reading, telling stories, and naming, counting, and drawing as the infant in their charge ages because they are more likely to have access to books, paper, and other implements that facilitate these didactic activities. Another factor that may contribute to differences in didactic activities in low- versus high-HDI countries is the greater prevalence of maternal depression in lower income countries (Hanlon, 2013).

In contrast to these cognitive kinds of caregiving, the patterns of relations between caregivers playing with, singing to, and taking infants out of doors and infant age were nonlinear. That is, the percentages of caregivers who engaged in these more socioemotional activities were not only greater but followed a different overall developmental trajectory: Like didactic caregiving, they increased incrementally from birth through 4 to 5 months of age, but from 6 through 11 months socioemotional caregiving functions decelerated or flattened. Thus, the percentages of caregivers who played, sang, and took their infants outside essentially gelled by the middle of the 1st year. This pattern was supported in 19–26 of 38 developing countries, including high-, medium-, and low-HDI groups with 84% (32 of 38) of countries showing nonlinear patterns in at least one kind of socioemotional caregiving. In a nutshell, more and more caregivers engage in playing, singing, and taking the infant in their charge out of doors as infants steadily increase in their socioemotional capabilities but only until around the midpoint of the 1st year.

Domains of caregiving practices with infants therefore generally followed one of two contrasting growth trajectories as revealed by cross-sectional microgenetic analysis. Differences in developmental functions for different kinds of caregiving could reflect caregivers' ethnotheories about the appropriateness of different kinds of caregiving for infants of different ages or infants' receptiveness or abilities to benefit from different experiences. Caregivers likely do not perceive newborns as so open to either didactic or socioemotional overtures as they do older infants, and only slowly grow in both these practices across the 1st half-year. At this point, growth functions diverge. Alternatively, in LMIC, which are beset by relatively high rates of infant mortality and morbidity, caregivers might only slowly display their increasing investment until infant survival is ensured (Lancy, 2013; Scheper-Hughes, 1989).

At around 6 months, attachments between infant and caregiver have likely formed (Bowlby, 1969) and mother and infant alike enjoy much more flexibility in the ways they respond to one another (Sroufe, 1997). Infants older than 6 months may therefore benefit equally from mutual cooperation and joint attention between them and their caregivers. By contrast, as they grow more and more infants are perceived to profit more from didactic caregiving. Mothers are thought to lead in exchanges with infants because they are the more mature or advanced partner in the dyad (Kochanska & Aksan, 2004; Vygotsky, 1930/1978), but infants play an active role in eliciting caregiving practices that are meaningful and relevant to their developmental needs (Bornstein, Tamis-LeMonda, Hahn, & Haynes, 2008). Infant caregivers look for and respond to particular infant cues that then elicit different kinds of interactions.

Gender

In answer to the fourth question about whether caregiving practices are the same or different for infant girls and boys, we found that developmental functions for all caregiving practices were similar for infants of the two genders. This finding may seem somewhat surprising. Parents are generally thought to hold different cognitions, beliefs, and expectations for their girls and boys, just as they are thought to treat them differently, even when infants do not differ in salient characteristics by their gender (Mondschein, Adolph, & Tamis-Le-Monda, 2000; Rubin, Provenzano, & Luria, 1974). However, differential treatment of young girls and boys is not necessarily a rule, and wide-ranging data from LMIC show caregivers often treat girls and boys quite similarly (Bornstein & Putnick, in press; Deater-Deckard & Lansford, in press). In some instances, infant gender per se reputedly encourages differential caregiving (Pomerleau, Bolduc, Malcuit, & Cossette, 1990). The absence of gender differences in the trajectories of caregiving of infant girls and boys might indicate that caregivers in LMIC possess ethnotheories that prompt similar treatment of female and male infants. Gender effects might be specific to the Western, educated, industrialized, rich, and democratic societies where gender has been studied (Henrich, Heine, & Norenzayan, 2010). Alternatively, girls and boys in LMIC might be reared similarly in infancy and differential gender treatment begins later in childhood. So much is fragile in the 1st year of life for children in LMIC that cultural proclivities pertaining to gender may not strongly manifest themselves until survival is more assured. Yet another possibility is that gender-differentiated treatment of infants occurs in more subtle or nuanced ways than captured by the MICS. For example, caregivers may take both girls and boys of doors, but they may be more likely to take one out gender or the other to locations that will somehow better that child's chances in life.

Caveats and Future Directions

This study has limitations that raise additional questions about infant caregiving in LMIC. The MICS relies on reports of caregiving activities. In some ways observations of actual caregiver practices might constitute a stronger database than self-report, but in other ways not; after all, caregiver reports constitute the prime social milieu of early childrearing (Bornstein, 2014b). It is useful to tally the different ways mothers, fathers, and others stimulate infants, and MICS items are helpful indicators. However, the MICS asked about a limited number of specific, if presumably universal (etic), caregiving items, and MICS items are dichotomous. Although they may be representative of infant caregiving, six MICS items should not be mistaken for the full spectrum of infant caregiving. Caregivers in LMIC engage in many other specific forms of caregiving, and they may privilege domains of development not tightly connected to the forms of caregiving captured by the MICS (Johnson & Adams, 2004).

Fully understanding caregiving practices and their meaning requires situating them in context (Bornstein, 1995). The same caregiving practice can have the same or different meanings in different contexts, just as different caregiving practices can have the same or different meanings in different contexts. European American parents use questions during joint book reading as a way to engage children (van Kleeck, 2003), where Tongans rely on recitation (McNaughton, 1995). In some locales, caregivers could display socioemotional involvement predominately through singing to a child, whereas in others caregivers could demonstrate affection physically. These different displays may serve the same socioemotional functions for caregivers in their respective societies. In addition, next to quantitative aspects of caregiving, qualitative aspects matter. Thus, two caregivers could equate in their engagement of play with infants, yet one might solicit sequences of high-level play that challenge and advance her infant's skills, whereas another might demonstrate low-level play that does not advance her infant's skills. Beyond considering the form and level of caregiving, it is critical also to consider the timing, content, and meaning of caregiving with respect to infants' ongoing activities.

Microgenetic analysis reveals that discovering and adopting a strategy can be slow or rapid. Here, we focused on uptake and expression of caregiving practices. A comprehensive description of rates of change requires information regarding both discovery and uptake. In microgenetic analysis, the so-called "risk set" includes participants who have not yet experienced an event of interest or, in the case here, have not yet adopted or expressed a new practice. This set decreases with experience or over time, as more participants discover or take up a practice. Understanding individual differences and determinants of the risk set versus those converted will be of future interest.

On a practical note, the prevalence rates of caregiving practices that do not require caregiver literacy or the availability of costly materials exceeded those that did. Thus, LMIC caregivers might be alerted to the fact that important child capabilities benefit from "free" naming and storytelling kinds of caregiving early in life. Ideally, all caregivers would engage in each of the six caregiving practices because each is known to provide infants with vital life experiences that stimulate developing brains and foster emerging cognitive and socioemotional skills.

Conclusions

Child survival is achieved through caregiver protection and nurturance, but child thriving is fostered through caregiving that involves sharing information through education and inculcating interpersonal competencies through socialization. Caregivers who so engage infants in their charge also gain access to their infants' emotional competence, social style, and cognitive potential, and they learn about their infants' proclivities, capabilities, and limits. Such knowledge can lead to more appropriate and beneficial interactions with the salutary result of enhanced child development and well-being.

References

Abelson, R. (1985). A variance explanation paradox: When a little is a lot. *Psychological Bulletin*, *97*, 129–133. doi:10.1037/0033-2909.97.1.129

Bornstein, M. H. (1995). Form and function: Implications for studies of culture and human development. *Culture and Psychology*, *1*, 123–137. doi:10.1177/1354067X9511009

Bornstein, M. H. (2006). Parenting science and practice. In W. Damon (Series Ed.) & K. A. Renninger &

I. E. Sigel (Vol. Eds.), *Handbook of child psychology: Vol. 4. Child psychology in practice* (6th ed., pp. 893–949). New York, NY: Wiley.

Bornstein, M. H. (2013). Parenting × Gender × Culture × Time. In W. B. Wilcox & K. K. Kline (Eds.), *Gender and parenthood: Biological and social scientific perspectives* (pp. 91–119). New York, NY: Columbia University Press.

Bornstein, M. H. (2014a). Human infancy . . . and the rest of the lifespan. *Annual Review of Psychology, 65*, 121–158. doi:10.1146/annurev-psych-120710-100359

Bornstein, M. H. (2014b). Parents' reports about their children's lives. In G. B. Melton, A. Ben-Arieh, J. Cashmore, G. S. Goodman, & N. K. Worley (Eds.), *The SAGE handbook of child research* (pp. 486–533). London, UK: Sage.

Bornstein, M. H. (2015). Children's parents. In R. M. Lerner Series (Ed.) & M. H. Bornstein & T. Leventhal (Vol. Eds.), *Handbook of child psychology and developmental science: Vol. 4. Ecological settings and processes in developmental systems* (7th ed., pp. 55–132). Hoboken, NJ: Wiley.

Bornstein, M. H., Arterberry, M. E., & Lamb, M. E. (2014). *Development in infancy: A contemporary introduction* (5th ed.). New York, NY: Psychology Press.

Bornstein, M. H., Cote, L. R., Haynes, O. M., Suwalsky, J. T. D., & Bakeman, R. (2012). Modalities of infant-mother interaction in Japanese, Japanese American immigrant, and European American dyads. *Child Development, 83*, 2073–2088. doi:10.1111/j.1467-8624.2012.01822.x

Bornstein, M. H., & Putnick, D. L. (in press). IV. Mothering and fathering daughters and sons in low- and middle-income countries. In M. H. Bornstein, D. L. Putnick, J. E. Lansford, K. Deater-Deckard, & R. H. Bradley (Eds.), *Gender in low- and middle-income countries. Monograph of the Society for Research in Child Development*. 81(1).

Bornstein, M. H., Putnick, D. L., Lansford, J. E., Deater-Deckard, K., & Bradley, R. (in press). Gender in low- and middle-income countries. *Monograph of the Society for Research in Child Development* 81(1).

Bornstein, M. H., Rebello Britto, P., Nonoyama-Tarumi, Y., Ota, Y., Petrovic, O., & Putnick, D. L. (2012). Child development in developing countries: Introduction and methods. *Child Development, 83*, 16–31. doi:10.1111/j.1467-8624.2011.01671.x

Bornstein, M. H., & Ruddy, M. G. (1984). Infant attention and maternal stimulation: Prediction of cognitive and linguistic development in singletons and twins. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance X: Control of language processes* (pp. 433–445). London, UK: Erlbaum.

Bornstein, M. H., Tamis-LeMonda, C., Hahn, C., & Haynes, O. M. (2008). Maternal responsiveness to very young children at three ages: Longitudinal analysis of a multidimensional modular and specific parenting construct. *Developmental Psychology, 44*, 867–874. doi:10.1037/0012-1649.44.3.867

Bowlby, J. (1969). *Attachment and loss: Attachment* (Vol. 1, 2nd ed.). New York, NY: Basic Books.

Bradley, R. H., Corwyn, R. F., McAdoo, H. P., & Garcia Coll, C. (2001). The home environments of children in the United States, Part 1: Variations by age, ethnicity, and poverty status. *Child Development, 72*, 1844–1867. doi:10.1111/1467-8624.t01-1-00382

Bronfenbrenner, U., & Morris, P. A. (2006). The bioecological model of human development. In R. M. Lerner & W. Damon (Eds.), *Handbook of child psychology: Vol. 1. Theoretical models of human development* (6th ed., pp. 793–828). Hoboken, NJ: Wiley.

Conger, R. D., & Donnellan, M. B. (2007). An interactionist perspective on the socioeconomic context of human development. *Annual Review of Psychology, 58*, 175–199. doi:10.1146/annurev.psych.58.110405.085551

de Rosnay, M., & Murray, L. (2012). Maternal care as the central environmental variable. In L. C. Mayes & M. Lewis (Eds.), *The Cambridge handbook of environment in human development* (pp. 58–81). New York, NY: Cambridge University Press.

De Wolff, M. S., & van IJzendoorn, M. H. (1997). Sensitivity and attachment: A meta-analysis on parental antecedents of infant attachment. *Child Development, 68*, 571–591. doi:10.1111/j.1467-8624.1997.tb04218.x

Deater-Deckard, K., & Lansford, J. E. (in press). V. Daughters' and sons' exposure to caregiving discipline and violence in low- and middle-income countries. In M. H. Bornstein, D. L. Putnick, J. E. Lansford, K. Deater-Deckard, & R. H. Bradley (Eds.), *Gender in low- and middle-income countries. Monograph of the Society for Research in Child Development*. 81(1).

Dercon, S., & Krishnan, P. (2009). Poverty and the psychosocial competencies of children: Evidence from the Young Lives sample in four developing countries. *Children, Youth and Environments, 19*, 138–163.

Edin, K., & Lein, L. (1997). *Making ends meet*. New York, NY: Russell Sage Foundation.

Emde, R. N. (1992). Social referencing research. In S. Feinman (Ed.), *Social referencing and the social construction of reality in infancy* (pp. 79–94). New York, NY: Springer.

Emde, R. N., Gaensbauer, T. J., & Harmon, R. J. (1976). *Emotional expression in infancy: A biobehavioral study*. Psychological Issues Monograph Series (Vol. 10, Monograph No. 37). New York, NY: International Universities Press.

Erikson, E. H. (1950). *Childhood and society*. New York, NY: Norton.

Freud, S. (1966). *Introductory lectures on psychoanalysis* (J. Strachey, Trans. & Ed.). New York, NY: Norton. (Original work published 1916–1917)

Göncü, A., & Gaskins, S. (Eds.). (2007). *Play and development: Evolutionary, sociocultural and functional perspectives*. Mahwah, NJ: Erlbaum.

Grantham-McGregor, S., Cheung, Y. B., Cueto, S., Glewwe, P., Richter, L., & Strupp, B., & the International Child Development Steering Group. (2007). Developmental potential in the first 5 years for children in developing countries. *Lancet, 369*, 60–70. doi:10.1016/S0140-6736(07)60032-4

Hanlon, C. (2013). Maternal depression in low- and middle-income countries. *International Health, 5*, 4–5. doi:10.1093/inthealth/ihs003

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33*, 61–135. doi:10.1017/S0140525X0999152X

Hoff, E., Laursen, B., & Tardif, T. (2002). Socioeconomic status and parenting. In M. H. Bornstein (Ed.), *Handbook of parenting: Vol. 2. Biology and ecology of parenting* (2nd ed., pp. 231–252). Mahwah, NJ: Erlbaum.

Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist, 60*, 581–592. doi:10.1037/0003-066X.60.6.581

Hyde, J. S. (2014). Gender similarities and differences. *Annual Review of Psychology, 65*, 373–398. doi:10.1037/0033-2909.132.1.33

Johnson, R. K., & Adams, A. V. (2004). *Skills development in Sub-Saharan Africa*. Washington, DC: World Bank.

Kochanska, G., & Aksan, N. (2004). Development of mutual responsiveness between parents and their young children. *Child Development, 75*, 1657–1676. doi:10.1111/j.1467-8624.2004.00808.x

Kolb, B., & Gibb, R. (2011). Brain plasticity and behavior in the developing brain. *Journal of the Canadian Academy of Child and Adolescent Psychiatry, 20*, 265–276.

Kuo, A. A., Franke, T. M., Regalado, M., & Halfon, N. (2004). Parent report of reading to young children. *Pediatrics, 113*, 1944–1951.

Lancy, D. F. (2013). "Babies aren't persons": A survey of delayed personhood. In H. Keller & O. Hiltrud (Eds.), *Different faces of attachment: Cultural variations of a universal human need* (pp. 66–112). Cambridge, UK: Cambridge University Press.

Leinaweaver, J. (2014). Informal kinship-based fostering around the world: Anthropological findings. *Child Development Perspectives, 8*, 131–136. doi:10.1111/cdep.12075

Leventhal, T., Dupéré, V., & Shuey, E. A. (2015). Children in neighborhoods. In R. M. Lerner (Series Ed.) & M. H. Bornstein & T. Leventhal (Vol. Eds.), *Handbook of child psychology and developmental science: Vol. 4. Ecological settings and processes in developmental systems* (7th ed., pp. 493–533). Hoboken, NJ: Wiley.

Lytton, H., & Gallagher, L. (2002). Parenting twins and the genetics of parenting. In M. H. Bornstein (Ed.), *Handbook of parenting: Children and parenting* (Vol. 1, 2nd ed., pp. 227–253). Mahwah, NJ: Erlbaum.

Lytton, H., & Romney, D. M. (1991). Parents' differential socialization of boys and girls: A meta-analysis. *Psychological Bulletin, 109*, 267–296. doi:10.1037/0033-2909.109.2.267

Maccoby, E. E., & Jacklin, C. N. (1974). *The psychology of sex differences* (Vol. 1). Stanford, CA: Stanford University Press.

Magnuson, K. A., & Duncan, G. J. (2002). Parents in poverty. In M. H. Bornstein (Ed.), *Handbook of parenting: Applied parenting* (Vol. 4, 2nd ed., pp. 95–121). Mahwah, NJ: Erlbaum.

Marsh, L. C., & Cormier, D. R. (2002). *Spline regression models* (Sage University Papers Series on Quantitative Applications in the Social Sciences, 07-137). Thousand Oaks, CA: Sage.

McCall, R. B. (1979). Qualitative transitions in behavioral development in the first two years of life. In M. Bornstein & W. Kessen (Eds.), *Psychological development from infancy* (pp. 183–224). Hillsdale, NJ: Erlbaum.

McNaughton, S. (1995). *Patterns of emergent literacy: Processes of development and transition*. Melbourne, Canada: Oxford University Press.

Mondschein, E. R., Adolph, K. E., & Tamis-LeMonda, C. S. (2000). Gender bias in mothers' expectations about infant crawling. *Journal of Experimental Child Psychology, 77*, 304–316. doi:10.1006/jecp.2000.2597

Morrison, C., & Jütting, J. P. (2005). Women's discrimination in developing countries: A new data set for better policies. *World Development, 33*, 1065–1081. doi:10.1016/j.worlddev.2005.04.002

Norenzayan, A., & Heine, S. J. (2005). Psychological universals: What are they and how can we know? *Psychological Bulletin, 131*, 763–784. doi:10.1037/0033-2909.131.5.763

Olson, S. L., Bates, J. E., & Bayles, K. (1984). Mother-infant interaction and the development of individual differences in children's cognitive competence. *Developmental Psychology, 20*, 166–179. doi:10.1037/0012-1649.20.1.166

Piaget, J. (1952). *The origins of intelligence in children*. New York, NY: Norton.

Pomerleau, A., Bolduc, D., Malcuit, G., & Cossette, L. (1990). Pink or blue: Environmental gender stereotypes in the first two years of life. *Sex Roles, 22*, 359–367. doi:10.1007/BF00288339

Power, T. G. (2000). *Play and exploration in children and animals*. Mahwah, NJ: Erlbaum.

Prentice, D. A., & Miller, D. T. (1992). When small effects are impressive. *Psychological Bulletin, 112*, 160–164. doi:10.1037/0033-2909.112.1.160

Quiroz, B. G., Snow, C. E., & Zhao, J. (2010). Vocabulary skills of Spanish-English bilinguals: Impact of mother-child language interactions and home language and literacy support. *International Journal of Bilingualism, 14*, 379–399. doi:10.1177/1367006910370919

Reddy, V. (2008). *How infants know minds*. Cambridge, MA: Harvard University Press.

Robinson, S. R., Adolph, K. E., & Young, J. W. (2004). *Continuity versus discontinuity: How different rime scales of behavioral measurement affect the pattern of developmental change*. Poster presented at the International Conference on Infancy Studies, Chicago, IL.

Rubin, J. Z., Provenzano, F. J., & Luria, Z. (1974). Eye of the beholder: Parents' views on the sex of newborns. *American Journal of Orthopsychiatry, 44*, 512–519. doi:10.1111/j.1939-0025.1974.tb00905.x

Sander, L. (1962). Issues in early mother-child interaction. *Journal of the American Academy of Child Psychiatry, 1*, 141–166. doi:10.1016/S0002-7138(09)60013-3

Scheper-Hughes, N. (1989, October). Human strategy: Death without weeping. *Natural History Magazine*, October, 8–16.

Siegler, R. S. (2006). Microgenetic analyses of learning. In W. Damon (Series Ed.) & D. Kuhn & R. S. Siegler (Vol. Eds.), *Handbook of child psychology: Vol. 2. Cognition, perception, and language* (6th ed., pp. 464–510). Hoboken, NJ: Wiley.

Smith, C. B., Adamson, L. B., & Bakeman, R. (1988). Interactional predictors of early language. *First Language, 8*, 143–156. doi:10.1177/014272378800802304

Smith, P. K., & Drew, L. M. (2002). Grandparenting. In M. H. Bornstein (Ed.), *Handbook of parenting: Status and social conditions of parenting* (Vol. 3, 2nd ed., pp. 141–172). Mahwah, NJ: Erlbaum.

Sroufe, L. A. (1996). *Emotional development: The organization of emotional life in the early years*. Cambridge, UK: Cambridge University Press.

Sroufe, L. A. (1997). Psychopathology as an outcome of development. *Development and Psychopathology, 9*, 251–268. doi:10.1017/S0954579497002046

Stern, D. (1985). *The interpersonal world of the infant*. New York, NY: Basic Books.

Tomlinson, M., Bornstein, M. H., Marlow, M., & Swartz, L. (2014). Imbalances in the knowledge about infant mental health in rich and poor countries: Too little progress in bridging the gap. *Infant Mental Health Journal, 35*, 624–629. doi:10.1002/imhj.21462

Trevarthen, C., & Hubley, P. (1978). Secondary intersubjectivity: Confidence, confiding and acts of meaning in the first year. In A. Lock (Ed.), *Action, gesture and symbol* (pp. 183–229). London, UK: Academic Press.

United Nations Children's Fund. (2006a). *Multiple indicator cluster survey manual 2005: Monitoring the situation of children and women*. New York, NY: Author.

United Nations Children's Fund. (2006b). *The state of the world's children 2007: Women and children The double dividend of gender equality*. New York, NY: Author.

United Nations Development Programme. (2008). *The human development report 2007/2008*. New York, NY: Palgrave Macmillan. Retrieved from http://hdr.undp.org/sites/default/files/reports/268/hdr_20072008_en_complete.pdf

United Nations Development Programme. (2011). *Human development report 2011. Sustainability and equity: A better future for all*. New York, NY: Palgrave Macmillan. Retrieved from http://hdr.undp.org/sites/default/files/reports/271/hdr_2011_en_complete.pdf

van Kleeck, A. (2003). Research on book sharing: Another critical look. In A. van Kleeck, S. A. Stahl, & E. B. Bauer (Eds.), *On reading books to children* (pp. 271–320). Mahwah, NJ: Erlbaum.

Vygotsky, L. S. (1962). *Thought and language*. New York, NY: Wiley. (Original work published 1934)

Vygotsky, L. S. (1978). *Mind and society: The development of higher mental processes*. Cambridge, MA: Harvard University Press (Original work published 1930).

Walker, S. P., Wachs, T. D., Gardner, J. M., Lozoff, B., Wasserman, G. A., & Pollitt, E., & the International Child Development Steering Group. (2007). Child development: Risk factors for adverse outcomes in developing countries. *Lancet, 369*, 145–157. doi:10.1016/S0140-6736(07)60076-2

Winnicott, D. W. (1964). *The child, the family and the outside world*. Harmondsworth, UK: Penguin Books.

World Bank. (2014). *How we classify countries*. Retrieved from http://data.worldbank.org/about/country-classifications

Child Development, September/October 2015, Volume 86, Number 5, Pages 1588–1603

Early Parenting and the Development of Externalizing Behavior Problems: Longitudinal Mediation Through Children's Executive Function

Michael J. Sulik and Clancy Blair
New York University

Roger Mills-Koonce
University of North Carolina at Greensboro

Daniel Berry
University of Illinois, Urbana-Champaign

Mark Greenberg
The Pennsylvania State University

The Family Life Project Investigators

Path analysis was used to investigate the longitudinal associations among parenting and children's executive function and externalizing behavior problems from 36 to 90 months of age in the Family Life Project (N = 1,115), a study of child development in the context of rural poverty. While controlling for stability in the constructs, semistructured observations of parenting prospectively predicted performance on a battery of executive function tasks and primary caregivers' reports of externalizing behavior. Furthermore, the association between early parenting and later externalizing behavior was longitudinally mediated by executive function, providing support for a process model in which sensitive parenting promotes children's self-regulation, which in turn reduces children's externalizing behavior.

Externalizing behavior problems such as aggression are relatively frequent in young children and, when persistent, are associated with elevated risk for psychopathology in adulthood (Conduct Problems Prevention Research Group, 1992). Executive function, an aspect of self-regulation encompassing the "attention shifting, working memory, and inhibitory control cognitive processes utilized in planning, problem solving, and goal-directed activity" (Blair & Razza, 2007, p. 648), has been identified as a robust correlate of child externalizing behavior problems in cross-sectional studies (Ogilvie, Stewart, Chan, & Shum, 2011). Nonetheless, there are few longitudinal studies relating these two constructs, so the direction of the association between executive function and externalizing behavior is somewhat unclear. Strong executive function could be used to support the regulation of externalizing behavior or negative emotions associated with externalizing behavior such as anger. It seems plausible, however, that these behavioral patterns and emotions could also interfere with children's ability to effectively deploy executive function, perhaps slowing its development over time (Bridgett et al., 2009; Leve et al., 2013). Indeed, such associations may well be bidirectional (Blair & Raver, 2012). Furthermore, parenting has been identified as an important predictor of both executive function and externalizing behavior, and could potentially influence the development of both of these constructs.

In this article, we investigate the longitudinal associations among parenting, child executive function, and child externalizing behavior from 36 to 90 months of age using data from the Family Life Project, a prospective longitudinal study of child development in the context of rural poverty. We use mediation analysis to test a process model of development in which sensitive parenting promotes chil-

---

The Family Life Project Phase II Key investigators include: Lynne Vernon-Feagans, The University of North Carolina at Chapel Hill; Mark T. Greenberg, The Pennsylvania State University; Clancy B. Blair, New York University; Margaret R. Burchinal, The University of North Carolina at Chapel Hill; Martha Cox, The University of North Carolina at Chapel Hill; Patricia T. Garrett-Peters, The University of North Carolina at Chapel Hill; Jennifer L. Frank, The Pennsylvania State University; W. Roger Mills-Koonce, University of North Carolina-Greensboro; and Michael T. Willoughby, RTI International.

We would like to thank the many families and research assistants who made this study possible. Support for this research was provided by the National Institute of Child Health and Human Development Grants R01 HD51502 and P01 HD39667 with cofunding from the National Institute on Drug Abuse.

Correspondence concerning this article should be addressed to Michael J. Sulik, 196 Mercer St., Rm. 810, New York, NY 10012. Electronic mail may be sent to michael.sulik@nyu.edu.

© 2015 The Authors
Child Development © 2015 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2015/8605-0019
DOI: 10.1111/cdev.12386

dren's capacity for self-regulation, which in turn reduces children's externalizing behavior. In addition, we investigate the possibility of bidirectional associations between parenting and children's executive function and externalizing behavior.

Parenting: Associations With Externalizing Symptoms

In young children, there is already substantial rank-order stability in externalizing behavior over a period of several years (Calkins, Blandon, Williford, & Keane, 2007). Despite this high stability (which limits researchers' ability to predict change), parenting consistently predicts children's externalizing behavior in prospective longitudinal studies that control for earlier externalizing behavior (Combs-Ronto, Olson, Lunkenheimer, & Sameroff, 2009; Eisenberg et al., 2005; Lengua, 2006; Spinrad et al., 2007). However, few studies have tested reciprocal associations between externalizing behavior and parenting (with the exception of disciplinary practices; e.g., Lansford et al., 2011). Although children's externalizing behavior has been shown to predict parents' immediate responses (Patterson, 1982), longitudinal studies testing longer term associations between children's externalizing behavior and parenting in early childhood have produced inconsistent results: Some investigators have reported null findings (Eisenberg, Spinrad, Eggum, et al., 2010), whereas others have reported effects of child externalizing behavior on parenting while controlling for rank-order stability in the constructs (Combs-Ronto et al., 2009).

Parenting: Associations With Self-Regulation

There appears to be a bidirectional pattern of associations between parenting and children's self-regulation. Research has indicated that parenting predicts change in both executive function (Hammond, Müller, Carpendale, Bibok, & Liebermann-Finestone, 2012) and effortful control (Kochanska, Murray, & Harlan, 2000; Lengua, Honorado, & Bush, 2007), which is an aspect of temperamental self-regulation that overlaps with executive function (Zhou, Chen, & Main, 2012). The reverse direction of effect, from self-regulation to subsequent parenting, has also been documented in preschool- and early, school-age children. For example, in an analysis using data from Blair, Raver, Berry, and The Family Life Project Investigators (2014) reported that executive function at 36 months predicted change in parenting from 36 to 60 months. Similarly, using data from the National Institute of Child Health and Human Development (NICHD) Study of Early Child Care and Youth Development, Belsky, Fearon, and Bell (2007) reported that executive attention predicted change in parenting from 54 months to first grade and from fourth grade to fifth grade (but not from first grade to third grade). However, null associations were found in community samples when effortful control was tested as a predictor of parenting from 18 to 30 months of age and from 30 to 42 months of age (Eisenberg, Spinrad, Eggum, et al., 2010) and in older samples (Eisenberg et al., 2005; Valiente et al., 2006).

Self-Regulation: Associations With Externalizing Symptoms

Meta-analysis (Ogilvie et al., 2011) has indicated that there is a moderate association between executive function and externalizing behavior, and that controlling for attention deficit hyperactivity disorder symptoms reduces—but does not eliminate—this association. There have been few longitudinal, prospective studies, however, that have examined the associations between executive function and externalizing behavior. Riggs, Blair, and Greenberg (2004) reported that executive function prospectively predicted externalizing behavior 2 years later while controlling for stability. Similarly, Nigg, Quamma, Greenberg, and Kusche (1999) reported that one aspect of executive function (inhibitory control; i.e., Color Word Stroop), but not a second (visual attention and task switching; i.e., Trails B), was prospectively related to teachers' reports of externalizing behavior 2 years later while controlling for stability in externalizing behavior. Comparable longitudinal studies relating effortful control to externalizing behavior have been more common and generally indicate that effortful control is negatively related to change in externalizing behavior after 54 months of age but that the evidence for longitudinal associations in younger children is equivocal (for a review of these studies, see Eisenberg, Spinrad, & Eggum, 2010). Little is known about the alternative direction of effect (externalizing behavior → executive function) in younger children, but externalizing behavior has been found to inconsistently predict effortful control across a 1-year lag in early childhood (Eisenberg, Spinrad, Eggum, et al., 2010), and null associations between externalizing behavior and later executive attention have been reported in middle childhood (Belsky et al., 2007). Eisenberg, Spinrad, Eggum, et al. (2010) speculated that children with externalizing behavior problems may have fewer opportunities to practice and improve self-regulatory skills. Although there is limited theory and empirical evidence that externalizing behavior longitudinally predicts self-regulation, we did not want to prematurely foreclose on the alternative direction of prediction from externalizing behavior to executive function and the possibility of bidirectional or transactional pathways.

Mediation

Although no studies have examined children's executive function as a longitudinal mediator between parenting and externalizing behavior, several longitudinal studies have tested effortful control as a mediator of this association. Support for this mediational model in early childhood has been mixed. For example, Eisenberg, Spinrad, Eggum, et al. (2010) reported that effortful control (measured using questionnaires and laboratory tasks) at 30 months failed to mediate the association between 18-month parenting and 42-month externalizing behavior. In a somewhat comparable study of children of alcoholic parents, the longitudinal association between parenting at age 2 and externalizing behavior at age 5 was mediated by self-regulation at age 3, a construct that included (but was not limited to) a battery of effortful control tasks (Eiden, Edwards, & Leonard, 2007). Findings are more consistent at older ages, with results from several studies of older children (Belsky et al., 2007; Valiente et al., 2006) and adolescents (Eisenberg et al., 2005) indicating that the association between parenting and externalizing behavior is longitudinally mediated by effortful control.

The Present Investigation

Our goal was to examine the longitudinal relations among parenting, executive function, and externalizing behavior during a developmental period of rapid growth in children's self-regulation in a sample at risk for poor self-regulation and high externalizing behavior. Based on prior empirical findings from the effortful control literature, we hypothesized a longitudinal mediation model in which early parenting would show a negative indirect association with children's development of externalizing behavior, such that parenting would predict executive function, which in turn would predict externalizing behavior (parenting → executive function → externalizing behavior). A strength of our longitudinal panel design is that it allows us to specify and test alternative models, such as a bidirectional model in which child characteristics predict later parenting and an alternative mediation model in which parenting predicts child externalizing behavior, which in turn predicts child executive function (parenting → externalizing behavior → executive function).

Previous studies investigating similar mediation hypotheses have been limited by the use of questionnaire methods to assess multiple constructs (or have relied on a single, specific measure of self-regulation; see Belsky et al., 2007). When the same informant (e.g., parent, teacher) reports on multiple constructs, the presence of method effects can bias the associations among the constructs. In this study, questionnaires were used to assess children's externalizing behavior, whereas laboratory measures were used to assess parenting behavior and children's executive function. Furthermore, executive function was assessed using a battery of tasks to obtain a broad measure that included working memory, inhibitory control, and set shifting.

Method

Participants

We use data from the Family Life Project, a prospective longitudinal study of 1,292 children born between fall 2003 and fall 2004 in six rural counties in Pennsylvania and North Carolina. Low-income families were over-sampled in both states and African American families were over-sampled in North Carolina. Extensive details about recruitment procedures and sample characteristics have been published elsewhere (Vernon-Feagans, Cox, & The Family Life Project Key Investigators, 2013). Children were assessed at home at approximately 2, 6, 15, 24, 36, 48, 60, and 90 months of age (summer following first grade). At each of these time points, families were visited by two highly trained data collectors and parents completed questionnaires on family demographics, household characteristics, and children's behavior. At the 6-, 15-, 24-, 36-, and 60-month assessments, parents engaged in a semi-structured interaction with their child. Children were administered a battery of executive function tasks at the 36-, 48-, and 60-month assessments.

Measures

Executive Function

The executive function battery consisted of five tasks at the 36-month assessment and six tasks at the 48- and 60-month assessments. At the 48- and

60-month assessments, the executive function battery included three inhibitory control tasks, two working memory tasks, and one attention shifting task. The inhibitory control tasks included a Simon-like spatial conflict task, a Stroop-like silly sounds task, and a farm animal go/no-go task. The working memory tasks included a span-like task and a self-ordered pointing task. The attention shifting task was an item selection task modeled on the Dimensional Change Card Sort task (Willoughby, Wirth, & Blair, 2011). At the 36-month assessment, a similar (but not identical) spatial conflict task was used and the self-ordered pointing task was not administered.

Each task was presented in an open spiral-bound flipbook with pages that measured 8 in. × 14 in. For each task, research assistants established that the child knew colors and numbers and administered training trials and up to three practice trials if needed. If children failed to demonstrate an understanding of the goals of the task following the practice trials, the examiner discontinued that task. Full details regarding the administration rules, psychometric properties, and scoring approach for each of these tasks have been presented elsewhere (Willoughby, Wirth, et al., 2011).

Item response theory was used to generate expected a posteriori (EAP) scores for each task and these scores were equated over time using a calibration sample. For details about the generation of the EAP scores and information about the longitudinal measurement invariance of these scores, see Willoughby, Wirth, Blair, and The Family Life Project Investigators (2012). Within each assessment, a one-factor model fit the data well and a two-factor model did not provide better fit to the data (Willoughby, Blair, Wirth, Greenberg, & The Family Life Project Investigators, 2010; Willoughby et al., 2012). Within each assessment, EAP scores were averaged to form a composite measure of executive function ability. As is typical of executive function measures (Willoughby, Holochwost, Blanton, & Blair, 2014), reliability coefficients for the composites at each time point were relatively low, $\alpha s = .37$ at 36 months ($n = 973$), .55 at 48 months ($n = 1{,}008$), and .50 at 60 months ($n = 1{,}038$).

Externalizing Behavior

Primary caregivers (typically biological mothers) completed the five-item Conduct Problems (e.g., "Often fights with other children or bullies them") scale from the Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997) at the 36- ($n = 1{,}093$), 48- ($n = 1{,}062$), 60- ($n = 1{,}084$), and 90- ($n = 1{,}088$) month assessments. The Conduct Problems scale is broadly representative of externalizing behavior problems. The SDQ items have three response categories (0 = *not true*, 1 = *somewhat true*, 2 = *certainly true*). Because Cronbach's α underestimates reliability for ordinal scales (especially for ordinal scales with few response categories), we report ordinal α reliability coefficients for these scales (Zumbo, Gaderman, & Zeisser, 2007). Ordinal αs were .76, .78, .81, and .87 at the 36-, 48-, 60-, and 90-month assessments, respectively.

Parenting

Children and their primary caregiver completed a 10-min semistructured free-play task at the 6- and 15-month assessments and a 10-min puzzle task at the 24- and 36-month assessments. The free-play interaction involved asking each caregiver to use a standardized set of toys and to play with the infant as he or she normally would. The puzzle task presented the child with a series of three increasingly difficult jigsaw puzzles to complete and asking the primary caregiver to assist the child in any way that he or she chose. At the 60-month assessment, primary caregivers and children participated in two consecutive tasks for a combined 12-min period. The first task was a tower construction task in which they were asked to recreate a tower structure using smaller wooden blocks of different shapes. The second task was a "slap jack" card game during which each player took turns putting down a card and slapped the pile whenever a jack card was played (to win the cards in the pile). These parent–child interactions were video recorded, and the following seven aspects of parenting were later coded for each of the 6- through 36-month assessments: sensitivity, detachment, intrusiveness, stimulation, positive regard, negative regard, and animation in interacting with the child (Cox & Crnic, 2002; NICHD Early Child Care Research Network, 1999). At the 60-month assessment, the same parenting scales were coded as at previous assessments except that respect for child autonomy replaced the intrusiveness code and the animation code was dropped. Ratings were made on a scale ranging from 1 (*not at all characteristic*) to 5 (*highly characteristic*) at the 6- and 15-month assessments and on a scale ranging from 1 (*not at all characteristic*) to 7 (*highly characteristic*) at the 24-, 36-, and 60-month assessments; to maintain consistency across time, scores at the latter visits were rescaled to range from 1 to 5.

Each scale was coded by a team of 4–5 coders, which included 1–2 master coders. Each coder was trained to be reliable with the master coder(s) and each coder completed approximately 30% of the tapes with the master coder(s). Although the composition of the coding teams changed over time, at least one of the master coders was consistent from one assessment to the next (e.g., a master coder who participated at the 6-month assessment also participated at the 15-month assessment). Intraclass correlations for independent ratings were used to assess reliability. For each of the parenting scales, the intraclass correlation exceeded .80 at each assessment.

Factor analysis of the parenting variables yielded two dimensions of parenting behavior from 6 to 36 months (Mills-Koonce et al., 2011; Vernon-Feagans, Cox, & The Family Life Project Key Investigators, 2013). Based on those results, we utilize a composite measure of parenting that includes the sensitivity, detachment (reversed), stimulation, positive regard, and animation codes. Intrusiveness and negative regard loaded on the second parenting factor. To simplify the presentation of results, we only include the first parenting factor in our analyses; substituting the second parenting factor did not substantively change the results. Cronbach's αs for the five-item parenting composite were .89, .90, .87, and .80 at the 6-, 15-, 24-, and 36-month assessments. We averaged the parenting scores across these four assessments to create a measure of early parenting quality, $\alpha = .86$, which for simplicity we subsequently refer to as "36-month parenting." At the 60-month assessment, we averaged the sensitivity, detachment (reversed), stimulation, and positive regard variables to create a similar measure of parenting quality, $\alpha = .83$.

Demographic Covariates

State of residence (Pennsylvania = 0, North Carolina = 1) was included as a covariate to control for site differences in study variables. At the 2-month assessment, primary caregivers reported on their number of years of education and the gender (0 = male, 1 = female) and race (0 = not African American, 1 = African American) of their child. In addition, primary caregivers and, when applicable, secondary caregivers provided information about household income and the number of individuals in the household at the 6-, 15-, 24-, and 36-month assessments. The household income-to-needs ratio was calculated as household income divided by the concurrent U.S. federal poverty threshold for a household of that size (U.S. Department of Health and Human Services, 2014), such that a value of 1.0 corresponds to household income equal to the poverty threshold (e.g., $18,850 for a family of four in 2004). These scores were highly correlated across time, with *r*s ranging from .73 to .85. Household income-to-needs was therefore averaged across the 6- through 36-month assessments, $\alpha = .94$, $M = 1.86$, $SD = 1.53$. This variable was natural log transformed to normalize its distribution, which exhibited high skewness and kurtosis prior to transformation.

Results

Missing Data

The full sample of the Family Life Project was composed of 1,292 participants at the 2-month assessment. Complete data were available for state of residence, gender and race of the child, and primary caregiver years of education. The percentages of missing data for other study variables were as follows: 36-month income-to-needs = 5.5%, 36-month executive function = 24.7%, 48-month executive function = 22.0%, 60-month executive function = 19.7%, 36-month externalizing behavior = 15.4%, 48-month externalizing behavior = 17.8%, 60-month externalizing behavior = 17.8%, 90-month externalizing behavior = 16.1%, 36-month parenting = 5.5%, and 60-month parenting = 25.5%.

We used SAS 9.3 and a macro obtained from http://www.appliedmissingdata.com/macro-programs.html to compute Little's (1988) test for data missing completely at random, $\chi^2(df = 605) = 888.354$, $p < .001$. This result indicates that the pattern of missing data is systematic (i.e., associated with measured or unmeasured variables). To assess potential bias due to differential attrition, we correlated missing data indicators (0 = not missing, 1 = missing) for each study variable with the demographic covariates and early parenting. Although there were a number of modest but significant correlations between these two sets of variables, none of these correlations exceeded .13.

Participants were included in the analytic sample ($N = 1,115$), which was used for all subsequent analyses, if they had nonmissing data at one or more assessments for executive function, externalizing behavior, and parenting. All models were specified and fitted in Mplus 7.11 (Muthén & Muthén, 2012) using the full information maximum likelihood estimator.

Table 1
Correlations and Descriptive Statistics

		1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	North Carolina	—													
2	Female	**0.07**	—												
3	African American	**0.63**	−0.01	—											
4	Caregiver education	**−0.17**	−0.01	**−0.22**	—										
5	Ln(income-to-needs)	**−0.29**	−0.04	**−0.44**	**0.60**	—									
6	36-month parenting	**−0.30**	0.02	**−0.42**	**0.51**	**0.53**	—								
7	60-month parenting	**−0.27**	0.01	**−0.39**	**0.47**	**0.51**	**0.70**	—							
8	36-month EF	**−0.29**	**0.10**	**−0.29**	**0.20**	**0.26**	**0.30**	**0.28**	—						
9	48-month EF	**−0.31**	**0.14**	**−0.34**	**0.31**	**0.32**	**0.39**	**0.34**	**0.38**	—					
10	60-month EF	**−0.22**	**0.13**	**−0.28**	**0.26**	**0.29**	**0.37**	**0.30**	**0.33**	**0.60**	—				
11	36-month EXT	**0.09**	−0.06	**0.14**	**−0.27**	**−0.30**	**−0.33**	**−0.26**	**−0.14**	**−0.19**	**−0.23**	—			
12	48-month EXT	**0.07**	**−0.07**	**0.13**	**−0.26**	**−0.25**	**−0.28**	**−0.22**	**−0.16**	**−0.21**	**−0.22**	**0.54**	—		
13	60-month EXT	0.03	−0.05	**0.12**	**−0.27**	**−0.26**	**−0.26**	**−0.26**	−0.09	**−0.21**	**−0.25**	**0.51**	**0.55**	—	
14	90-month EXT	−0.03	**−0.09**	0.05	**−0.22**	**−0.21**	**−0.23**	**−0.17**	**−0.10**	**−0.15**	**−0.25**	**0.49**	**0.54**	**0.55**	—
M		0.59	0.43	0.50	14.45	0.94	2.86	2.88	−0.56	−0.13	0.29	0.63	0.44	0.39	0.38
SD		—	—	—	2.80	0.44	0.66	0.76	0.55	0.51	0.48	0.41	0.34	0.36	0.39

Note. EF = executive function; EXT = externalizing behavior. Bold values indicate $p < .05$.

Descriptive Statistics and Correlations

Descriptive statistics and correlations for all study variables are presented in Table 1. On average, primary caregivers reported 14.45 years of education (*SD* = 2.80 years), and 19.8% of primary caregivers reported fewer than 12 years of education. Prior to transformation, the average household income-to-needs ratio was 1.86 (*SD* = 1.53), and 67.8% of participating households' average income from 6 to 36 months was below 200% of the poverty threshold. To characterize the severity of externalizing behavior in the Family Life Project sample, we compared the means on the conduct problems scale from the 48-, 60-, and 90-month assessments to gender-matched national norms for the United States from the National Health Interview Survey for children aged 4–7 (Youthinmind Ltd., 2004). As expected based on the high-risk nature of the sample, Family Life Project participants scored higher than the age 4–7 norms at all three assessments, average Cohen's *d*s = 0.43 for girls and 0.32 for boys.

There was considerable rank-order stability in parenting from 36 to 60 months, *r* = .70. There was also rank-order stability for all child constructs that were assessed multiple times, with cross-time correlations ranging from .33 to .60 for executive function and from .49 to .55 for externalizing behavior. Executive function was negatively related to concurrently measured externalizing behavior at the 36-, 48-, and 60-month assessments, *r*s = −.14, −.21, and −.25. Parenting at 36 and 60 months was positively related to executive function and negatively related to externalizing behavior across all time points, with *r*s ranging from .28 to .39 for executive function and from −.17 to −.33 for externalizing behavior.

Longitudinal Panel Model

In a preliminary analysis step, we estimated autoregressive models for executive function and externalizing behavior separately. Using likelihood ratio tests, models in which each score was predicted by the scores at all previous time points fit significantly better than simpler models in which all longitudinal paths were not included. Because the more complex, completely saturated autoregressive models showed better fit to the data and because controlling for all prior scores provides a more conservative test of cross-lagged paths in a panel model with multiple constructs, we retained the more complex autoregressive models for use in subsequent analyses.

We combined the autoregressive models for executive function and externalizing behavior described above and added demographic variables (i.e., state, gender, race, primary caregiver education, and family income-to-needs) at the initial assessment and parenting at 36 and 60 months. All within-time correlations among exogenous variables were estimated at the 36-month assessment, and standardized residual covariances among concurrently measured exogenous variables were estimated at subsequent assessments. To maintain a parsimonious model, we only included longitudinal paths for demographic covariates when indicated by substantial changes in model fit: (a) executive function at 48 months was regressed on gender, (b) executive function at 48 months was regressed on African American race, and (c) parenting at 60 months was regressed on the family income-to-needs ratio. Furthermore, the inclusion of these paths did not substantively change the results. This model—which is depicted (without covariates) in Figure 1—fit the data well, $\chi^2(df = 37) = 189.064$, $p < .001$, root mean square error of approximation

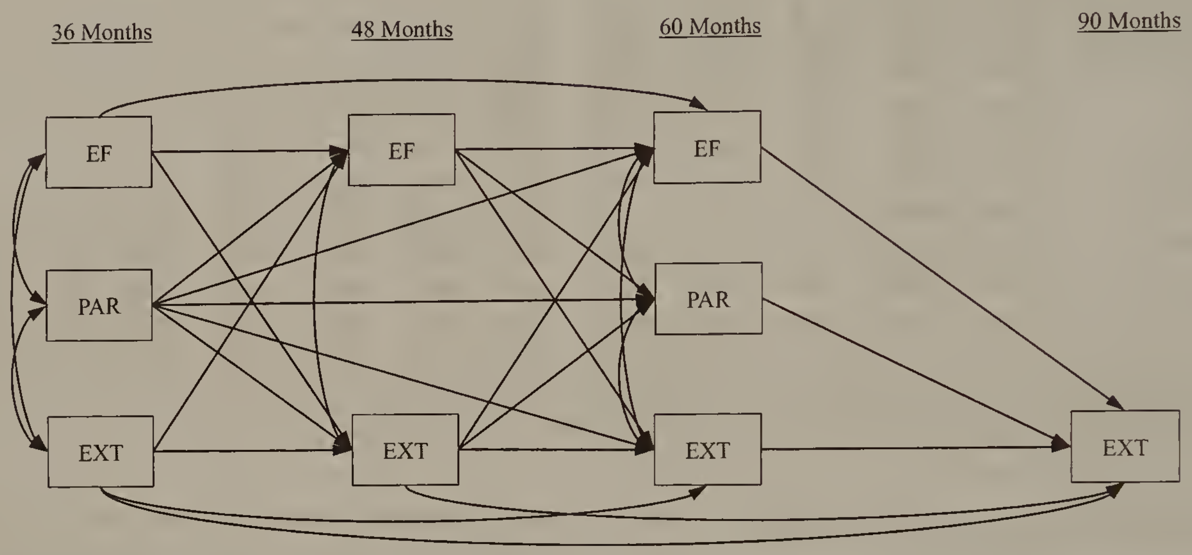


Figure 1. Hypothesized autoregressive model with cross-lagged paths. EF = executive function; EXT = externalizing behavior; PAR = parenting. Our statistical model included covariates that are not depicted in this figure.

Table 2
Directional Paths in the Longitudinal Panel Model

	Dependent variable								
	Executive function			Externalizing			Parenting		
Predictor	*b*	*SE*	β	*b*	*SE*	β	*b*	*SE*	β
				48 months					
Parenting 36	0.182***	0.025	.232	−0.048**	0.015	−.093			
EF 36	0.233***	0.029	.248	−0.040*	0.018	−.065			
EXT 36	−0.058	0.036	−.047	0.408***	0.023	.502			
				60 months					
Parenting 36	0.095***	0.020	.130	−0.023	0.015	−.042	0.669***	0.032	.579
EF 36	0.085**	0.027	.096						
EF 48	0.464***	0.027	.493	−0.045*	0.019	−.065	0.077*	0.038	.052
EXT 36				0.239***	0.026	.276			
EXT 48	−0.104**	0.037	−.073	0.405***	0.032	.381	−0.006	0.054	−.002
				90 months					
Parenting 60				0.017	0.014	.034			
EF 60				−0.063**	0.021	−.078			
EXT 36				0.173***	0.029	.186			
EXT 48				0.299***	0.036	.261			
EXT 60				0.325***	0.033	.302			

Note. Empty cells indicate that no directional path was estimated. EF = executive function; EXT = externalizing behavior.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

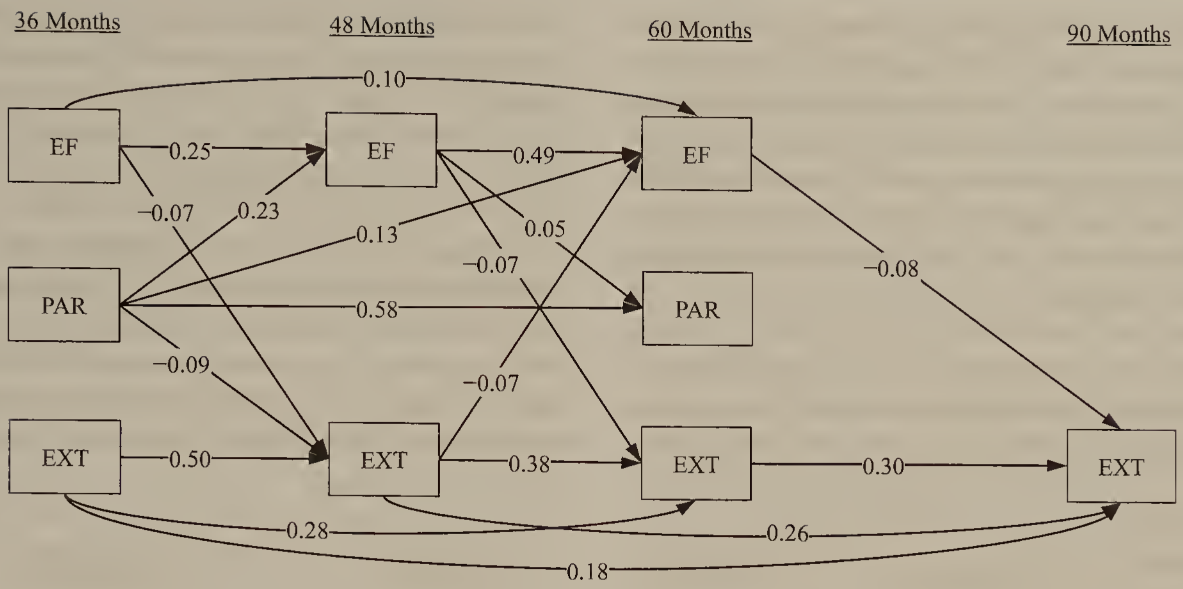


Figure 2. Significant directional paths in autoregressive model with cross-lagged paths. EF = executive function; EXT = externalizing behavior; PAR = parenting. Only significant directional paths (with standardized βs) are shown. Our statistical model included nonsignificant directional paths, covariates, correlations among all exogenous variables at the first assessment, and correlations among residuals within each subsequent assessment that are not depicted in this figure.

(RMSEA) = 0.061, 90% CI [0.052, 0.069], comparative fit index (CFI) = 0.969, standardized root mean square residual (SRMR) = 0.038.

The standardized and unstandardized coefficients for all directional paths among the executive function, externalizing behavior, and parenting variables in this model (with standard errors) are presented in Table 2 and the statistically significant directional paths in the model are also depicted (with standardized βs) in Figure 2. As expected, we found evidence for an association between early parenting and later executive function and externalizing behavior: Parenting at 36 months predicted executive function at both 48 months and 60 months, βs = .248 and .130, and was also predictive of externalizing behavior at 48 months (but not of externalizing behavior at 60 months), β = −.093. Contrary to expectations, parenting at 60 months was unrelated to externalizing behavior at 90 months. There was evidence for modest prediction of parenting from child characteristics, with executive function (but not externalizing behavior) at 48 months predicting parenting at 60 months, β = .052. There was a small but significant standardized residual covariance between parenting and externalizing behavior at 60 months, φ = −.08, $p < .01$.

The cross-lagged paths between executive function and externalizing behavior allow us to address questions about the direction of influence between these two constructs. As expected based on prior research, there was consistent prediction from executive function to externalizing behavior at the subsequent assessment: Executive function at 36 months predicted externalizing behavior at 48 months, executive function at 48 months predicted externalizing behavior at 60 months, and executive function at 60 months predicted externalizing behavior at 90 months, βs = −.065, −.065, and −.078. The opposite direction of prediction—from externalizing behavior to executive function at the subsequent assessment—was significant at 48 to 60 months, β = −.073, but not from 36 to 48 months. There were modest standardized residual covariances between executive function and externalizing behavior at 48 and 60 months, φs = −.07 and −.09, $p < .05$ and $p < .01$.

The R^2 values for the endogenous variables in this model were as follows: 48-month executive function = 26.6%, 60-month executive function = 39.1%, 48-month externalizing behavior = 30.7%, 60-month externalizing behavior = 37.6%, 90-month externalizing behavior = 40.9%, and 60-month parenting = 51.4%. To quantify the effect size of the direct effects of executive function, externalizing behavior, and parenting we fixed these paths to zero for each of these variables and computed the difference in R^2 between the constrained and unconstrained models. The direct effect of early parenting uniquely explained 3.5% of the variance in executive function at 48 months, 1.1% of the variance in executive function at 60 months, and 0.5% of the variance in externalizing behavior at 48 months, whereas executive function at 48 months explained only 0.1% of the variance in parenting at 60 months. Executive function at 36 months explained 0.4% of the variance in externalizing behavior at 48 months, executive function at 48 months explained 0.3% of the variance in externalizing behavior at 60 months, and executive function at 60 months explained 0.5% of the variance in externalizing behavior at 90 months. With respect to the opposite direction of prediction, externalizing behavior at 48 months explained 0.4% of the variance in executive function at 60 months.

To examine whether there were gender differences in the paths, we estimated a multigroup path model for boys and girls in which all covariances and directional paths were free to vary across groups. We then compared this model to a nested model in which all covariances and directional paths were constrained to be equal for boys and girls. The unconstrained model fit well, $\chi^2(df = 58) = 126.258$, $p < .001$, RMSEA = .046, 90% CI [.035, .057], CFI = .983, SRMR = .029, as did the constrained model, $\chi^2(df = 108) = 180.488$, $p < .001$, RMSEA = .035, 90% CI [.026, .043], CFI = .982, SRMR = .039. To compare models, we used a likelihood ratio test, which indicated that the constrained model did not fit worse relative to the

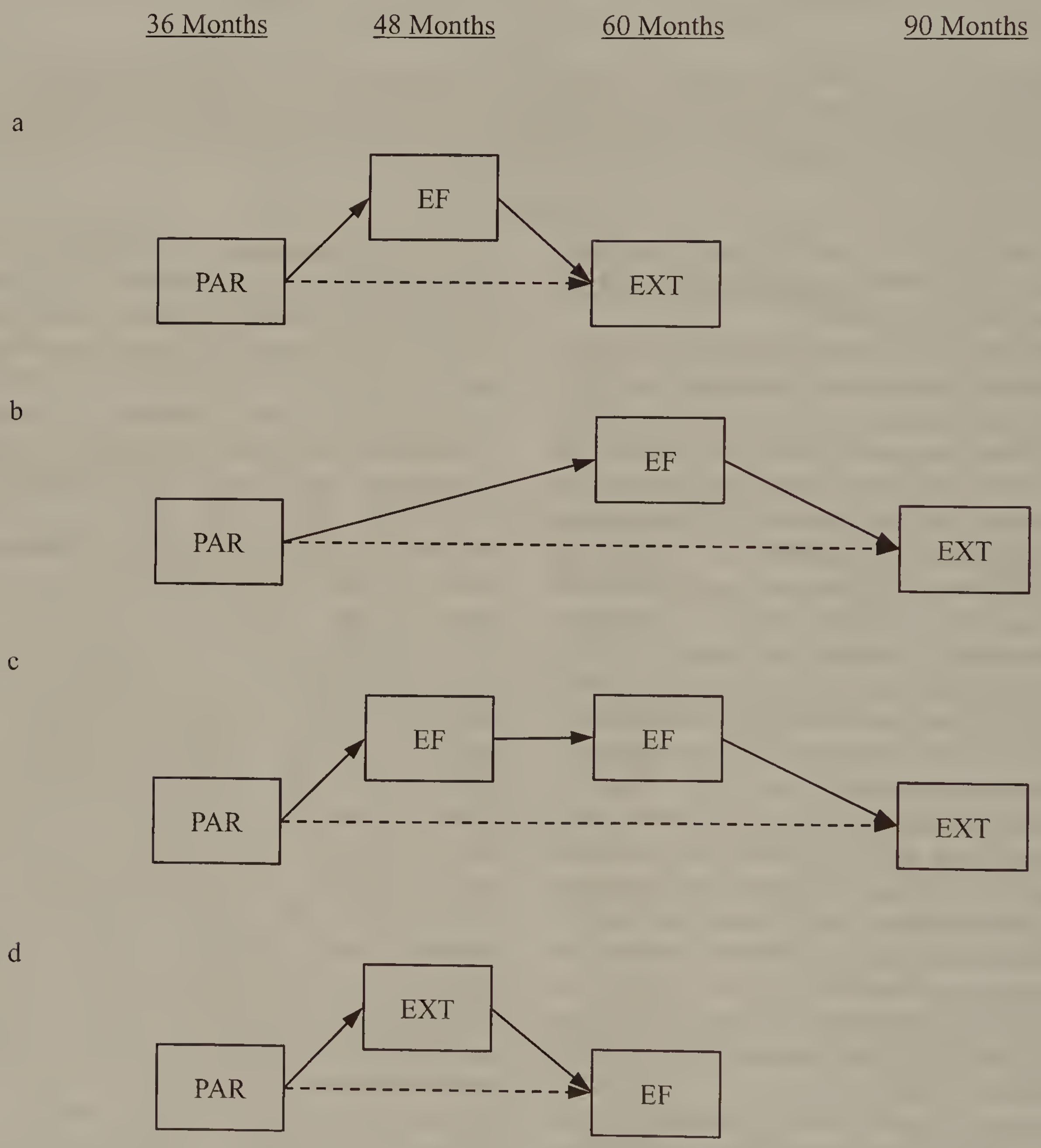


Figure 3. Mediation model examples. EF = executive function; EXT = externalizing behavior; PAR = parenting. Solid lines indicate the paths between the predictor and a mediator, between two mediators, or between a mediator and the criterion, whereas dashed lines indicate the direct effect of the predictor on the criterion while accounting for the indirect effects. Panel a = indirect effect of parenting on externalizing at 60 months through executive function at 48 months; Panel b = indirect effect of parenting on externalizing at 90 months through executive function at 60 months; Panel c = indirect effect of parenting on externalizing at 90 months through executive function at 48 and 60 months; Panel d = indirect effect of parenting on executive function at 60 months through externalizing at 48 months.

unconstrained model, $\chi^2(df = 50) = 54.230$, $p = .316$. As a result, we did not consider this multigroup model further.

Mediation

Our primary hypothesis was that the association between parenting and later externalizing behavior would be mediated by executive function. We therefore tested the indirect effect of 36-month parenting on 60-month externalizing behavior through 48-month executive function (see Figure 3a), and the indirect effect of 36-month parenting on 90-month externalizing behavior through 60-month executive function (see Figure 3b). In addition, we tested mediational pathways involving executive function that involved more than a single mediator (see Figure 3c). There were two such pathways of interest: (a) from 36-month parenting to 48-month executive function to 60-month executive function to 90-month externalizing behavior, and (b) from 36-month parenting to 48-month executive function to 60-month externalizing behavior to 90-month externalizing behavior. We were also interested in the effects of parenting on later externalizing behavior that were mediated by stability in externalizing behavior (e.g., from 36-month parenting to 48-month externalizing behavior to 60-month externalizing behavior).

In addition to testing the hypothesized model in which the effect of parenting on later externalizing behavior was mediated by executive function, we also evaluated an alternative model in which the effect of parenting on executive function would be mediated by externalizing behavior. Therefore, we also tested the indirect effect of 36-month parenting on 60-month executive function through 48-month externalizing (see Figure 3d).

All indirect effects were evaluated using the panel model that we previously presented, with one minor change: We added the (nonsignificant) direct path between 36-month parenting and 90-month externalizing behavior so that the total effect of 36-month parenting could be decomposed into direct and indirect effects. All indirect effects were tested using the percentile bootstrap with 5,000 bootstrap sample draws (Taylor, MacKinnon, & Tein, 2008). To quantify effect size, we report the completely standardized indirect effects (Preacher & Kelley, 2011) with bootstrapped 95% confidence

Table 3
Mediation Analysis: Direct and Indirect Effects of Early Parenting

	Dependent variable					
	Executive function			Externalizing behavior		
Predictor	*b*	*SE*	β	*b*	*SE*	β
60 months						
PAR (direct effect)	0.095***	0.020	.130	−0.023	0.016	−.042
Specific indirect effects						
PAR → EF 48	0.084***	0.013	.115	−0.008*	0.004	−.015
PAR → EXT 48	0.005*	0.003	.007	−0.019**	0.006	−.035
90 months						
PAR (direct effect)				−0.012	0.020	−.020
Specific indirect effects through EF						
PAR → EF 48 → EF 60				−0.005*	0.002	−.009
PAR → EF 48 → EXT 60				−0.003*	0.001	−.005
PAR → EF 60				−0.006*	0.002	−.010
Sum of specific indirect effects through EF				−0.013**	0.004	−.023
Specific indirect effects through EXT						
PAR → EXT48				−0.014**	0.005	−.024
PAR → EXT 48 → EXT 60				−0.006**	0.002	−.011
PAR → EXT 60				−0.007	0.005	−.013
Sum of specific indirect effects through EXT				−0.028**	0.008	−.047

Note. Specific indirect effects are the sum of the specific indirect effects listed in this table. There are additional, nonsignificant indirect effects in the model. PAR = parenting; EF = executive function; EXT = externalizing behavior.
$^*p < .05$. $^{**}p < .01$. $^{***}p < .001$.

intervals. The β coefficients for the indirect effects indicate how much the dependent variable would be expected to change (expressed on the metric of standard deviations) for a 1 *SD* change in the predictor.

First we tested the hypothesized (Figure 3a) and alternative mediation models (Figure 3d) using dependent variables at the 60-month assessment (see Table 3). As hypothesized, the association between 36-month parenting and 60-month externalizing behavior was mediated by 48-month executive function, β = −.015. However, there was also support for the alternative model in which the association between 36-month parenting and 60-month executive function was mediated by 48-month externalizing behavior, β = .007. Next, we tested the hypothesized model using the dependent variable at 90 months (see Table 3). Unfortunately, the alternative model could not be tested in this analysis because executive function data were not collected at the 90-month assessment. The effect of 36-month parenting on 90-month externalizing behavior was mediated by executive function at 60 months (see Figure 3b), β = −.010, and by two more complex pathways involving executive function at 48 months: one pathway from parenting to executive function at 48 months to executive function at 60 months (see Figure 3c), and one pathway from parenting to executive function at 48 months to externalizing behavior at 60 months (not depicted), βs = −.009 and −.005. The sum of these three indirect effects was negatively related to 90-month externalizing behavior, β = −.023. The effects of parenting were also carried forward through stability in executive function and in externalizing behavior. For example, 36-month parenting predicted 48-month executive function, which in turn predicted 60-month executive function. These indirect effects are also listed in Table 3.

Discussion

Our primary goal was to test a hypothesized longitudinal mediation model in which early parenting predicts child executive function, which in turn predicts child externalizing behavior problems (parenting → executive function → externalizing behavior). Because externalizing behavior was assessed four times, we were able to test the hypothesized mediation effect twice, and the hypothesized mediation model was supported in both of these instances: We found a significant, indirect effect of 36-month parenting on 60-month and 90-month externalizing behavior that operated through children's executive function at intermediate assessments. Although modest, the effect sizes should be interpreted in light of the high stability in all constructs, which limited our ability to predict across constructs. This is especially an issue for the effect size of the mediated effect, which is calculated as the product of two other coefficients. We calculated that the standardized βs for the indirect effects reported by Belsky et al. (2007) were −.044 and −.017. These values are comparable to the indirect effects for executive function that we reported in this study, βs = −.015 and −.023.

These mediation results provide evidence consistent with a causal model in which sensitive parenting reduces children's externalizing behavior primarily by promoting children's executive function, and parallels findings from a school-based intervention study in which executive function mediated the effect of intervention status on externalizing behavior (Riggs, Greenberg, Kusché, & Pentz, 2006). Critically, these results help inform our understanding of a potential, albeit weak, mechanism through which parenting might influence children's externalizing behavior problems. Executive function is believed to be a resource that children can use to flexibly regulate their behaviors and emotions and to plan. This volitional capacity for self-regulation could potentially be applied in a variety of ways to reduce externalizing behavior.

However, we are currently unable to rule out the possibility of the alternative mediation model (parenting → externalizing behavior → executive function), which also received some support. A transactional pattern of associations, in which executive function and externalizing behavior are mutually influential over time, remains a possibility.

Testing Bidirectional Relations

We were also able to investigate the associations between parenting and child executive function and externalizing behavior over time to test for bidirectional or transactional associations between parenting and child characteristics. While controlling for rank-order stability in the constructs, early parenting had a direct effect on executive function assessed 1 and 2 years later, and on externalizing behavior problems 1 year later. This pattern of associations mirrors prior work in which parenting has been reported to predict the subsequent development of children's self-regulation (Hammond et al., 2012). The effect size (for the direct effect) of early parenting on 48-month externalizing behavior was modest, $R^2 = 0.5\%$, but was more substantial for executive function at 48 months, $R^2 = 1.1\%$, and early parenting continued to independently predict executive function (but not externalizing behavior) at 60 months, $R^2 = 0.5\%$. Although relatively small, these R^2 values should be interpreted with respect to the high stability in the constructs, which limits the variability that can be explained by other predictors.

In contrast to the paths from parenting to child characteristics, there was only limited evidence of child characteristics predicting parenting: Child executive function at 48 months predicted parenting at 60 months, but this effect uniquely explained only 0.1% of the variance in 60-month parenting. For comparison, Blair et al. (2014) reported somewhat larger standardized regression coefficients using 36-month executive function as a predictor of 60-month parenting in the Family Life Project; however, their models did not include child externalizing behavior, which could potentially reduce the unique effect of executive function on parenting. The small effect size for the association between child executive function and parenting is in line with other studies finding weak or inconsistent evidence in support of child effects on parent behavior (Bates, Schermerhorn, & Petersen, 2012). It is possible, however, that so-called "coercive" child effects on parenting behavior (Patterson, 1982) are moderated by severity of externalizing behavior symptomatology, that these patterns have already been established early in life and therefore do not have additional effects later on parenting, or that the amount of time between assessments in this study was too long to detect such effects. Unfortunately, the absence of parenting data at the 48-month assessment also limited our ability to draw conclusions regarding child effects on parenting.

The longitudinal design also made it possible to examine the temporal ordering of the associations between executive function and externalizing behavior. Whereas executive function consistently predicted externalizing behavior at every subsequent assessment, the evidence for the alternative direction of prediction was more limited, with externalizing behavior predicting executive function at the subsequent assessment only one of two times. All effect sizes for the prediction across these constructs were modest. Consistent with Eisenberg, Spinrad, Eggum, et al. (2010) findings, externalizing behavior at 48 months was prospectively related to executive function at 60 months. At present, this evidence is somewhat inconclusive and, as previously stated, we cannot rule out a transactional developmental process in which executive function and externalizing behavior exert mutual influence on one another over time—or that an unmeasured confounding variable influences change in both executive function and externalizing behavior.

Effects for Early but Not Later Parenting

In contrast to the association between early parenting and subsequent externalizing behavior, parenting assessed at 60 months failed to predict externalizing behavior at 90 months while controlling for stability in externalizing behavior. This does not imply that parenting at 60 months and externalizing behavior at 90 months were unrelated. Rather, it means that we did not find additional prediction of children's externalizing behavior at 90 months from parenting at 60 months over and above prior externalizing behavior. Although the amount of time between assessments can affect the magnitude of the associations observed in longitudinal panel models (Selig & Little, 2012), the longer amount of time between the last two assessments (i.e., 30 months) than between the previous assessments (i.e., 12 months), cannot explain why early parenting predicts 90-month externalizing behavior, but 60-month parenting does not. It might be the case that parents regulate their children's behavior problems externally in early childhood, when children's self-regulatory abilities are weak, and that effective external regulation promotes effective self-regulation in their children that can then be used the manage externalizing behavior even in the absence of external regulation from caregivers (Eisenberg & Spinrad, 2004; Kopp, 1982). The stronger effect of early parenting relative to later parenting on 90-month externalizing behavior is consistent with theories of experiential canalization and developmental plasticity (Blair & Raver, 2012; Del Giudice, Ellis, & Shirtcliff, 2011) and suggests that parenting interventions targeting externalizing behavior would likely be most effective earlier, rather than later, in children's development.

Distinguishing Executive Function From Effortful Control

Although we focus on executive function in this study, prior research on effortful control has also indicated that the development of self-regulation is influenced by the early family environment. However, those studies have used questionnaires or relatively narrow assessments of self-regulation rather than a comprehensive battery of executive function tasks assessing working memory, shifting, and inhibition. Our study provides evidence that early parenting behavior is a robust predictor of subsequent executive function while controlling for stability (see also Blair et al., 2014), and a particular strength of our findings is that each of the three constructs (parenting, executive function, and externalizing behavior) was assessed using independent methods. Although externalizing behavior was assessed using only a single informant, the high mean levels are consistent with the high-risk nature of the Family Life Project sample and parent reports on the SDQ have been shown to be a valid screening measure for clinical levels of externalizing behavior problems (Goodman, 2001).

Batteries of laboratory tasks are frequently used in developmental studies to assess both executive function (Garon, Smith, & Bryson, 2014; Wiebe, Espy, & Charak, 2008) and effortful control (Kochanska & Knaack, 2003), and there is some overlap between the tasks that are used as measures of executive function and effortful control, with "day–night" (Gerstadt, Hong, & Diamond, 1994) and similar inhibition tasks being used as measures of both constructs. However, there is one major difference in the measurement of these two constructs: Effortful control batteries have not used tasks that assess primarily working memory or set shifting (Kochanska & Knaack, 2003), whereas these tasks are considered an important component of batteries assessing executive function (Garon et al., 2014; Miyake et al., 2000; Wiebe et al., 2008). In addition, effortful control batteries more commonly include delay tasks, which might differ from cognitive executive function tasks that lack a strong motivational component (Kim, Nordling, Yoon, Boldt, & Kochanska, 2012; Willoughby, Kupersmidt, Voegler-Lee, & Bryant, 2011). Furthermore, the use of questionnaire measures appears to be more common in the effortful control literature. At present, it is still unclear how much these differences in measurement contribute to differences in the associations between effortful control, executive function, and other constructs (Blair & Razza, 2007).

One small meta-analytic study has suggested that the effect size of the association between inhibitory control and externalizing behavior is greater than the association between other aspects of executive function (i.e., working memory; set shifting) and externalizing behavior (Schoemaker, Mulder, Deković, & Matthys, 2013). An important direction for future research will be to examine how the three foundational executive function abilities are differentially related to externalizing behavior in longitudinal studies.

Limitations, Conclusions, and Implications

The degree of rank-order stability in the constructs is negatively related to statistical power to detect prediction of change over time across constructs. When power is low, it is not clear whether the absence of an association is due to the failure to detect a true effect or the absence of a true association among the variables. Given high levels of stability for parenting and moderate levels of stability for externalizing behavior and executive function, a large sample size is therefore needed to adequately address questions of longitudinal mediation among these constructs. Our sample was ideally suited to address this research question because the large sample size affords much greater statistical power than most previous investigations (for an exception, refer to Belsky et al., 2007).

One limitation of our study was the relatively low reliability of the battery of executive function tasks, which was particularly acute at the initial assessment. In longitudinal mediation analyses, unreliability can have two undesirable and opposing effects. First, unreliability attenuates the estimates of the associations among the constructs and the degree of rank-order stability in each construct over time. This can potentially result in biased longitudinal prediction across constructs because variance in the variables at later assessments that *should* be accounted for by rank-order stability in the constructs and covariance among the constructs at an earlier assessment can potentially be predicted longitudinally. Second, unreliability in the mediating variable is known to negatively bias the indirect effect and positively bias the direct effect (Kenny, 2014), potentially leading to an underestimate of the true effect size of the mediated effect. However, the low reliability of the executive function measures is a limitation that is by no means unique to the Family Life Project, in which executive function was measured using a battery of 5–6 diverse executive function tasks with the goal of obtaining a comprehensive assessment of executive function ability.

Our model provides evidence that one way in which parents might influence the development of children's subsequent externalizing behavior is by scaffolding and nurturing their executive function skills (for a discussion of the mechanisms through which parenting might affect executive function, see Bernier, Carlson, & Whipple, 2010; Hughes & Ensor, 2009). However, we have limited knowledge about the specificity of parenting behavior and the development of children's executive function or externalizing behavior. In part, this is because different aspects of parenting tend to be interrelated, which makes it difficult to disentangle the effects of specific aspects of parenting. Our measure of parenting was reliable, longitudinally stable, and a conceptually broad measure, but it is unclear whether more specific aspects of parenting would provide better prediction than our broad measure of parenting quality.

References

Bates, J. E., Schermerhorn, A. C., & Petersen, I. T. (2012). Temperament and parenting in developmental perspective. In M. Zentner & R. L. Shiner (Eds.), *Handbook of temperament* (pp. 424–441). New York, NY: Guilford.

Belsky, J., Fearon, R. M., & Bell, B. (2007). Parenting, attention and externalizing problems: Testing mediation longitudinally, repeatedly and reciprocally. *Journal of Child Psychology and Psychiatry, 48,* 1233–1242. doi:10.1111/j.1469-7610.2007.01807.x

Bernier, A., Carlson, S. M., & Whipple, N. (2010). From external regulation to self-regulation: Early parenting precursors of young children's executive functioning. *Child Development, 81,* 326–339. doi:10.1111/j.1467-8624.2009.01397.x

Blair, C., & Raver, C. C. (2012). Child development in the context of adversity: Experiential canalization of brain and behavior. *American Psychologist, 67,* 309–318. doi:10.1037/a0027493

Blair, C., Raver, C. C., Berry, D. J., & The Family Life Project Investigators. (2014). Two approaches to estimating the effect of parenting on the development of executive function in early childhood. *Developmental Psychology, 50,* 554–565. doi:10.1037/a0033647

Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development, 78,* 647–663. doi:10.1111/j.1467-8624.2007.01019.x

Bridgett, D. J., Gartstein, M. A., Putnam, S. P., McKay, T., Iddins, E., Robertson, C., . . . Rittmueller, A. (2009). Maternal and contextual influences and the effect of temperament development during infancy on parenting in toddlerhood. *Infant Behavior and Development, 32,* 103–116. doi:10.1016/j.infbeh.2008.10.007

Calkins, S. D., Blandon, A. Y., Williford, A. P., & Keane, S. P. (2007). Biological, behavioral, and relational levels of resilience in the context of risk for early childhood behavior problems. *Development and Psychopathology, 19,* 675–700. doi:10.1017/S095457940700034X

Combs-Ronto, L. A., Olson, S. L., Lunkenheimer, E. S., & Sameroff, A. J. (2009). Interactions between maternal parenting and children's early disruptive behavior: Bidirectional associations across the transition from preschool to school entry. *Journal of Abnormal Child Psychology, 37,* 1151–1163. doi:10.1007/s10802-009-9332-2

Conduct Problems Prevention Research Group. (1992). A developmental and clinical model for the prevention of conduct disorder: The FAST Track Program. *Development and Psychopathology, 4,* 509. doi:10.1017/S0954579400004855

Cox, M., & Crnic, K. (2002). *Qualitative ratings for parent-child interaction at 3–12 months of age.* Unpublished manuscript, Department of Psychology, University of North Carolina at Chapel Hill.

Del Giudice, M., Ellis, B. J., & Shirtcliff, E. A. (2011). The adaptive calibration model of stress responsivity. *Neuroscience & Biobehavioral Reviews, 35,* 1562–1592. doi:10.1016/j.neubiorev.2010.11.007

Eiden, R. D., Edwards, E. P., & Leonard, K. E. (2007). A conceptual model for the development of externalizing behavior problems among kindergarten children of alcoholic families: Role of parenting and children's self-regulation. *Developmental Psychology, 43,* 1187–1201. doi:10.1037/0012-1649.43.5.1187

Eisenberg, N., & Spinrad, T. L. (2004). Emotion-related regulation: Sharpening the definition. *Child Development, 75,* 334–339. doi:10.1111/j.1467-8624.2004.00674.x

Eisenberg, N., Spinrad, T. L., & Eggum, N. D. (2010). Emotion-related self-regulation and its relation to children's maladjustment. *Annual Review of Clinical Psychology, 6,* 495–525. doi:10.1146/annurev.clinpsy.121208.131208

Eisenberg, N., Spinrad, T. L., Eggum, N. M., Silva, K. M., Reiser, M., Hofer, C., . . . Michalik, N. (2010). Relations among maternal socialization, effortful control, and maladjustment in early childhood. *Development and Psychopathology, 22,* 507–525. doi:10.1017/S0954579410000246

Eisenberg, N., Zhou, Q., Spinrad, T. L., Valiente, C., Fabes, R. A., & Liew, J. (2005). Relations among positive parenting, children's effortful control, and externalizing problems: A three-wave longitudinal study. *Child Development, 76,* 1055–1071. doi:10.1111/j.1467-8624.2005.00897.x

Garon, N., Smith, I. M., & Bryson, S. E. (2014). A novel executive function battery for preschoolers: Sensitivity to age differences. *Child Neuropsychology, 20,* 713–736. doi:10.1080/09297049.2013.857650

Gerstadt, C. L., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action: Performance of children 3.5–7 years old on a Stroop-like day-night test. *Cognition, 53,* 129–153. doi:10.1016/0010-0277(94)90068-X

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry, 38,* 581–586. doi:10.1111/j.1469-7610.1997.tb01545.x

Goodman, R. (2001). Psychometric properties of the Strengths and Difficulties Questionnaire. *Journal of the American Academy of Child & Adolescent Psychiatry, 40,* 1337–1345. doi:10.1097/00004583-200111000-00015

Hammond, S. I., Müller, U., Carpendale, J. I., Bibok, M. B., & Liebermann-Finestone, D. P. (2012). The effects of parental scaffolding on preschoolers' executive function. *Developmental Psychology, 48*, 271–281. doi:10.1037/a0025519

Hughes, C. H., & Ensor, R. A. (2009). How do families help or hinder the emergence of early executive function? *New Directions for Child and Adolescent Development, 2009*, 35–50. doi:10.1002/cd.234

Kenny, D. A. (2014, April 9). *Mediation*. Retrieved from http://davidakenny.net/cm/mediate.htm

Kim, S., Nordling, J. K., Yoon, J. E., Boldt, L. J., & Kochanska, G. (2012). Effortful control in "hot" and "cool" tasks differentially predicts children's behavior problems and academic performance. *Journal of Abnormal Child Psychology, 41*, 43–56. doi:10.1007/s10802-012-9661-4

Kochanska, G., & Knaack, A. (2003). Effortful control as a personality characteristic of young children: Antecedents, correlates, and consequences. *Journal of Personality, 71*, 1087–1112. doi:10.1111/1467-6494.7106008

Kochanska, G., Murray, K. T., & Harlan, E. T. (2000). Effortful control in early childhood: Continuity and change, antecedents, and implications for social development. *Developmental Psychology, 36*, 220–232. doi:10.1037//0012-1649.36.2.220

Kopp, C. B. (1982). Antecedents of self-regulation: A developmental perspective. *Developmental Psychology, 18*, 199–214. doi:10.1037/0012-1649.18.2.199

Lansford, J. E., Criss, M. M., Laird, R. D., Shaw, D. S., Pettit, G. S., Bates, J. E., & Dodge, K. A. (2011). Reciprocal relations between parents' physical discipline and children's externalizing behavior during middle childhood and adolescence. *Development and Psychopathology, 23*, 225–238. doi:10.1017/S0954579410000751

Lengua, L. J. (2006). Growth in temperament and parenting as predictors of adjustment during children's transition to adolescence. *Developmental Psychology, 42*, 819–832. doi:10.1037/0012-1649.42.5.819

Lengua, L. J., Honorado, E., & Bush, N. R. (2007). Contextual risk and parenting as predictors of effortful control and social competence in preschool children. *Journal of Applied Developmental Psychology, 28*, 40–55. doi:10.1016/j.appdev.2006.10.001

Leve, L. D., DeGarmo, D. S., Bridgett, D. J., Neiderhiser, J. M., Shaw, D. S., Harold, G. T., . . . Reiss, D. (2013). Using an adoption design to separate genetic, prenatal, and temperament influences on toddler executive function. *Developmental Psychology, 49*, 1045–1057. doi:10.1037/a0029390

Little, R. J. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association, 83*, 1198–1202. doi:10.1080/01621459.1988.10478722

Mills-Koonce, W. R., Garrett-Peters, P., Barnett, M., Granger, D. A., Blair, C., Cox, M. J., & The Family Life Project Key Investigators. (2011). Father contributions to cortisol responses in infancy and toddlerhood. *Developmental Psychology, 47*, 388–395. doi:10.1037/a0021066

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*, 49–100. doi:10.1006/cogp.1999.0734

Muthén, L. K., & Muthén, B. O. (2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén. Retrieved from http://www.statmodel.com

National Institute of Child Health and Human Development Early Child Care Research Network. (1999). Chronicity of maternal depressive symptoms, maternal sensitivity, and child functioning at 36 months. *Developmental Psychology, 35*, 1297–1310. doi:10.1037/0012-1649.35.5.1297

Nigg, J. T., Quamma, J. P., Greenberg, M. T., & Kusche, C. A. (1999). A two-year longitudinal study of neuropsychological and cognitive performance in relation to behavioral problems and competencies in elementary school children. *Journal of Abnormal Child Psychology, 27*, 51–63. doi:10.1023/A:1022614407893

Ogilvie, J. M., Stewart, A. L., Chan, R. C., & Shum, D. H. (2011). Neuropsychological measures of executive function and antisocial behavior: A meta-analysis. *Criminology, 49*, 1063–1107. doi:10.1111/j.1745-9125.2011.00252.x

Patterson, G. R. (1982). *Coercive family process*. Eugene, OR: Castalia.

Preacher, K. J., & Kelley, K. (2011). Effect size measures for mediation models: Quantitative strategies for communicating indirect effects. *Psychological Methods, 16*, 93–115. doi:10.1037/a0022658

Riggs, N. R., Blair, C. B., & Greenberg, M. T. (2004). Concurrent and 2-year longitudinal relations between executive function and the behavior of 1st and 2nd grade children. *Child Neuropsychology, 9*, 267–276. doi:10.1076/chin.9.4.267.23513

Riggs, N. R., Greenberg, M. T., Kusché, C. A., & Pentz, M. A. (2006). The mediational role of neurocognition in the behavioral outcomes of a social-emotional prevention program in elementary school students: Effects of the PATHS curriculum. *Prevention Science, 7*, 91–102. doi:10.1007/s11121-005-0022-1

Schoemaker, K., Mulder, H., Deković, M., & Matthys, W. (2013). Executive functions in preschool children with externalizing behavior problems: A meta-analysis. *Journal of Abnormal Child Psychology, 41*, 457–471. doi:10.1007/s10802-012-9684-x

Selig, J. P., & Little, T. D. (2012). Autoregressive and cross-lagged panel models for longitudinal data. In B. Laursen, T. D. Little, & N. A. Card (Eds.), *Handbook of developmental research methods* (pp. 265–290). New York, NY: Guilford.

Spinrad, T. L., Eisenberg, N., Gaertner, B., Popp, T., Smith, C. L., Kupfer, A., . . . Hofer, C. (2007). Relations of maternal socialization and toddlers' effortful control to children's adjustment and social competence. *Devel-*

opmental Psychology, 43, 1170–1186. doi:10.1037/0012-1649.43.5.1170

Taylor, A. B., MacKinnon, D. P., & Tein, J.-Y. (2008). Tests of the three-path mediated effect. *Organizational Research Methods, 11,* 241–269. doi:10.1177/1094428107300344

U.S. Department of Health and Human Services. (2014). *Prior HHS poverty guidelines and Federal Register references.* Retrieved from http://aspe.hhs.gov/poverty/figures-fed-reg.cfm

Valiente, C., Eisenberg, N., Spinrad, T. L., Reiser, M., Cumberland, A., Losoya, S. H., & Liew, J. (2006). Relations among mothers' expressivity, children's effortful control, and their problem behaviors: A four-year longitudinal study. *Emotion, 6,* 459–472. doi:10.1037/1528-3542.6.3.459

Vernon-Feagans, L., Cox, M., & The Family Life Project Key Investigators. (2013). The Family Life Project: An epdemiological and developmental study of young children living in poor rural communities. *Monographs of the Society for Research in Child Development, 78*(5).

Wiebe, S. A., Espy, K. A., & Charak, D. (2008). Using confirmatory factor analysis to understand executive control in preschool children: I. Latent structure. *Developmental Psychology, 44,* 575–587. doi:10.1037/0012-1649.44.2.575

Willoughby, M., Holochwost, S. J., Blanton, Z. E., & Blair, C. B. (2014). Executive functions: Formative versus reflective measurement. *Measurement: Interdisciplinary Research and Perspectives, 12,* 69–95. doi:10.1080/15366367.2014.929453

Willoughby, M. T., Blair, C. B., Wirth, R. J., & Greenberg, M.; The Family Life Project Investigators. (2010). The measurement of executive function at age 3 years: Psychometric properties and criterion validity of a new battery of tasks. *Psychological Assessment, 22,* 306–317. doi:10.1037/a0018708

Willoughby, M. T., Kupersmidt, J., Voegler-Lee, M., & Bryant, D. (2011). Contributions of hot and cool self-regulation to preschool disruptive behavior and academic achievement. *Developmental Neuropsychology, 36,* 162–180. doi:10.1080/87565641.2010.549980

Willoughby, M. T., Wirth, R. J., & Blair, C. B. (2011). Contributions of modern measurement theory to measuring executive function in early childhood: An empirical demonstration. *Journal of Experimental Child Psychology, 108,* 414–435. doi:10.1016/j.jecp.2010.04.007

Willoughby, M. T., Wirth, R. J., & Blair, C. B.; The Family Life Project Investigators. (2012). Executive function in early childhood: Longitudinal measurement invariance and developmental change. *Psychological Assessment, 24,* 418–431. doi:10.1037/a0025779

Youthinmind Ltd. (2004, January 30). *Normative SDQ data from the USA.* Retrieved from http://www.sdqinfo.org/USNorm.html

Zhou, Q., Chen, S. H., & Main, A. (2012). Commonalities and differences in the research on children's effortful control and executive function: A call for an integrated model of self-regulation. *Child Development Perspectives, 6,* 112–121. doi:10.1111/j.1750-8606.2011.00176.x

Zumbo, B. D., Gaderman, A. M., & Zeisser, C. (2007). Ordinal versions of coefficients alpha and theta for Likert rating scales. *Journal of Modern Applied Statistical Methods, 6,* 21–29.

Child Development, September/October 2015, Volume 86, Number 5, Pages 1604–1622

The Sexualized Girl: A Within-Gender Stereotype Among Elementary School Children

Ellen A. Stone, Christia Spears Brown, and Jennifer A. Jewell
University of Kentucky

Two studies (conducted in 2013) examined whether elementary-aged children endorse a within-gender stereotype about sexualized girls. In Study 1, children ($N = 208$) ages 6–11 rated sexualized girls as more popular but less intelligent, athletic, and nice compared to nonsexualized girls. These distinctions were stronger for girls and older children, and in accordance with our developmental intergroup theoretical framework, were related to children's cognitive development and media exposure. Study 2 ($N = 155$) replicated the previous findings using more ecologically valid and realistic images of girls and further explored individual differences in the endorsement of the sexualized girl stereotype. Additional results indicated that the belief that girls should be appearance focused predicted their endorsement of the sexualized girl stereotype.

The study of gender stereotypes among children largely began in the 1970s (e.g., Kuhn, Nash, & Brucken, 1978; Tibbetts, 1975; Williams, Bennett, & Best, 1975). Research since then has clearly documented that children endorse gender stereotypes as young as age 2 and continue to hold stereotyped beliefs throughout childhood (Albert & Porter, 1983; Golombok et al., 2008). Traditionally, most research has shown that girls are perceived as being kind, gentle, passive, sensitive, and well behaved relative to boys (Albert & Porter, 1983; Miller, Lurye, Zosuls, & Ruble, 2009), and these stereotypes have been reflected and perpetuated throughout children's media (Aubrey & Harrison, 2004; Browne, 1998; Hamilton, Anderson, Broaddus, & Young, 2006).

There has been a cultural shift, however, over the past two decades in how girls are portrayed in media and society at large; girls, even young girls, are increasingly being portrayed as sexualized (American Psychological Association, 2007). Sexualization occurs when women and girls are most valued for their sexual appeal, are sexually objectified, or have sexuality inappropriately forced on them (American Psychological Association, 2007). Although research has documented the increasing sexualization of girls, very little research has focused on how children perceive these sexualized girls, and whether children (both boys and girls) have distinct attitudes and stereotypes about sexualized girls relative to nonsexualized girls. This is the purpose of the current two studies.

The Sexualization of Girls

To date, most research on the sexualization of girls has documented the increasingly prevalent media portrayals of sexualized women and girls, and examined the impact of sexually objectifying messages on girls' and boys' attitudes about themselves and others. Before discussing previous research, it is important to clarify the different, but overlapping, constructs.

First, *sexualization* differs from a general *appearance orientation* (or being appearance focused). Appearance orientation involves self-surveillance and monitoring one's appearance, with a focus on being attractive, and is associated with being well groomed, clean, and "natural" (Smolak, Murnen, & Myers, 2014). Although both boys and girls can show concerns about their appearance, girls are typically more appearance focused than boys (Dunn, Lewis, & Patrick, 2010; Jones & Crawford, 2006). Indeed, a focus on appearance is a key component of the feminine gender norms that influence

---

The authors are grateful to Maya Kearns and the children and parents of Central Kentucky. Also, the authors especially thank Petra Ronald, as well as Jenny Baker, Allie Campbell, Camila Cardosa, Jordan Carr, Naomi Charalambakis, Meredith Haley, Chia Hong, Shelby Kennard, Bria Mills, Francesca Reynaert, Jonathan Sarfin, Ishman Sims, Abby Spalding, Casey Taylor, Adrian Weldon, and Rachel Williams.

Correspondence concerning this article should be addressed to Christia Spears Brown, Department of Psychology, University of Kentcuky, 125 Kastle Hall, Lexington, KY 40506. Electronic mail may be sent to christia.brown@uky.edu.

© 2015 The Authors
Child Development © 2015 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2015/8605-0020
DOI: 10.1111/cdev.12405

European American middle-class culture in the United States (Mahalik et al., 2005). Appearance orientation is contrasted with sexualization, which refers specifically to behaviors that relate to being "sexy," and is associated with wearing heavy makeup and tight clothing that emphasizes sexual body parts (e.g., breasts and buttocks; Smolak et al., 2014). A person can be attractive, but not necessarily sexualized, and girls are much more likely to be shown as sexualized than boys (Hatton & Trautner, 2011; Near, 2013). In the current studies, sexualization is operationalized as wearing tight clothing that emphasizes sexual body parts and wearing makeup (in accordance with Smolak et al., 2014).

Sexualization is also related to, albeit distinct from, *objectification* (defined as being turned into an object for someone else's use). Although it is possible to be sexualized without being objectified, they are often confounded and most of the research on sexualization has really examined the impact of the sexual objectification of women and girls. Specifically, research has shown that television, music videos, magazines, and video games are saturated with messages in which women and girls are portrayed as sexual objects for men's pleasure (Aubrey, Hopper, & Mbure, 2011; Conrad, Dixon, & Zhang, 2009; Daniels, 2009; Downs & Smith, 2010; Ward, 1995, 2002). Thus, women and girls are both sexualized, via their tight clothing and makeup, and objectified, portrayed as existing for men's pleasure and lacking in their own agency. There is not a parallel trend in which boys are often sexualized and objectified.

Because most portrayals of sexualized women and girls also involve objectification, research has consistently shown that repeated sexualization messages are associated with numerous negative outcomes (Fredrickson & Roberts, 1997; Lerum & Dworkin, 2009). For example, after watching images of sexually objectified women, girls show lower body esteem, greater dieting restrictions, lower self-esteem, and depression (American Psychological Association, 2007; Fredrickson, Roberts, Noll, Quinn, & Twenge, 1998; Glick, Larsen, Johnson, & Branstiter, 2005). Girls who internalize messages about sexualization have lower academic performance and more body shame than girls who do not (McKenney & Bigler, 2014a, 2014b). Although boys are not sexualized like girls are, they are still affected by these messages about girls; for boys, viewing sexually objectified girls is associated with rape myth acceptance, endorsement of the stereotype that women are sex objects and men are sexual pursuers, and endorsement of traditional gender stereotypes (American Psychological Association, 2007).

The current studies focus on the sexualization of girls. This is important to examine within developmental psychology because girls—not just women—are often sexualized, and both boys and girls are the targets of sexualized messages. Children's television shows frequently portray girls as sexualized by wearing tight, revealing clothing (Lacroix, 2004), an especially concerning trend given that children between the ages of 2 and 11 years old watch an average of 24 hrs of television a week (Turrill, 2013). Even children's clothing has sexualized characteristics, with one fourth of girls' clothing described as being revealing or having sexually suggestive writing (Goodin, Denburg, Murnen, & Smolak, 2011). Previous research has found that girls as young as 4 are already portraying sexualized behaviors, such as wearing makeup and having body image concerns (Tiggemann & Slater, 2014).

It is also important to examine stereotypes about sexualized girls within the context of gender stereotypes. Researchers have repeatedly documented the negative impact of gender stereotypes (Bigler, Arthur, Hughes, & Patterson, 2008; Else-Quest, Hyde, & Linn, 2010; Liben, Bigler, & Krogh, 2001). For example, adults and children often endorse the traditional gender stereotype that girls are bad at math, and this stereotype can be detrimental for girls (e.g., by reducing their math performance via stereotype threat, self-fulfilling prophecy, and expectancy-value mechanisms, as well as leading to academic discrimination; Leaper & Brown, 2014). Because of the repeated media portrayal of sexualized girls, there is also a stereotype that girls should be sexualized. This may be detrimental for girls, especially if the stereotype is associated with other negative traits. However, while most research on sexualization has examined the impact of watching sexualized media, little research has examined whether children (both boys and girls) spontaneously notice the sexualization of girls, whether children have specific stereotypes or beliefs about sexualized girls, and how these stereotypes about sexualized girls are related to their stereotypes about nonsexualized girls and their attitudes about girls in general.

While there has been little research on children's stereotypes, research has shown that adults have stereotypes about sexualized women and girls and that these stereotypes are indeed associated with negative traits. Sexualized women appear to be perceived as less competent, less intelligent, and less moral (albeit desired by men) compared to nonsexualized women (Daniels & Zurbriggen, 2014; Graff, Murnen, & Smolak, 2012; Grauerholz & King, 1997; Smolak et al., 2014; Ward, 1995). Furthermore, sexualized women are judged more on their appearance than their task performance (Glick et al., 2005; Rudman & Borgida, 1995). These negative traits are applied to sexualized girls as well. For example, men and women perceived a sexualized fifth-grade girl to be less determined, capable, and competent than a nonsexualized girl (Graff et al., 2012). Only one known study has been conducted, however, examining children's perceptions of sexualized girls. In a study by Starr and Ferguson (2012), researchers found that by age 9, girls perceive sexualized dolls to be more favorable and desirable than nonsexualized dolls, suggesting that children do distinguish between nonsexualized and sexualized girls. This research, however, did not include boys' attitudes about sexualized girls and did not examine children's specific stereotypes about sexualized girls.

Thus, the goal of the current studies was to examine children's stereotypes about sexualized girls. Identifying the existence of a "sexualized girl stereotype" will allow for a more systematic study of the impact of sexualization and may lead to interventions based on existing stereotyping interventions. We predicted that children would perceive sexualized girls differently than nonsexualized girls. Specifically, we predicted that children would have distinct stereotypes about sexualized girls that include her being more popular, but not as smart, athletic, or nice as nonsexualized girls. These traits were predicted to be associated with sexualized girls because (a) common media images of sexualized girls show them being popular, but not smart or nice (e.g., *Mean Girls*) and (b) the adult literature indicates that sexualized women are perceived as less intelligent, less competent, and less moral than nonsexualized women. In this way, the "sexualized girl stereotype" operates like a within-gender stereotype (or subtype) of a girl. Thus, using a developmental intergroup theoretical framework (Bigler & Liben, 2006), we examined whether boys and girls hold a within-gender stereotype of sexualized girls.

The Construction of Stereotypes: Developmental Intergroup Theory

Developmental intergroup theory details the process through which children develop stereotypes about particular groups (Bigler & Liben, 2006). Based on this framework, we propose that children first perceive sexualization to be a perceptually and psychologically salient attribute; they then categorize individuals based on that attribute; and finally, they construct a stereotype about sexualized girls that they apply to new individuals. We outline each step below.

Perceptual and Psychological Salience of Group Attributes

For children to hold a stereotype about a group, they must first perceive a group attribute to be perceptually and psychologically salient. In other words, children must be able to easily recognize and encode the group attribute before they develop stereotypes based on that group. Experimental research has shown that group attributes that are not highly salient (such as eye color) are not used as a basis for group stereotypes (Bigler, 1995). Salient group attributes, and groups for which children develop strong stereotypes, include gender, race, and attractiveness (Bigler & Liben, 2006). To examine whether children hold stereotypes about sexualized girls, we must first demonstrate that being sexualized is a perceptually and psychologically salient group attribute for children.

Categorization of Individuals

Second, according to developmental intergroup theory, if a group attribute is salient, children will categorize individuals by that salient dimension. The extent to which children perceive and categorize individuals based on their group attribute is dependent on their cognitive abilities (Bigler & Liben, 2006). One cognitive ability that relates to children's categorization is their classification skills. This refers to the ability to classify objects, animals, or people into discrete categories (e.g., the ability to sort dogs and cats). Multiple-classification ability refers to the ability to categorize individuals along more than one dimension simultaneously (e.g., brown and black dogs and cats), a skill that develops around age 7 (Piaget & Inhelder, 1969).

The attainment of these classification skills has been linked with the development of stereotypes (Aboud, 1989; Bigler, 1995). For example, multiple-classification abilities allow children to categorize an individual as a girl who can also be an athlete, which can lead to a decrease in between-gender stereotyping in children (Bigler & Liben, 1992). We predicted, however, that multiple-classification skills could also result in increased complexity in children's *within*-gender stereotypes, such that children would be able to categorize an individual as a specific type of girl (e.g., a nonsexualized girl may also be an athlete, but a sexualized girl cannot be an athlete). We predicted that children who have developed multiple-classification skill would be more likely to categorize sexualized girls as a subgroup of girls, in comparison to children who have not yet developed multiple-classification skill and who would be less likely to differentiate sexualized from nonsexualized girls.

Development of Stereotypes

Finally, according to our framework, once children perceive a group to be salient, and categorize individuals based on that group, they begin to develop stereotypes about the salient social group. Specifically, they make group–attribute links, often based on exposure to environmental (e.g., media) messages (Bigler & Liben, 2006). For example, because the media often portray sexualized girls as attractive and popular (Herbozo, Tantleff-Dunn, Gokee-Larose, & Thompson, 2004), children may link the group "sexualized girls" with the attribute of popularity.

In addition to linking salient groups with specific attributes, because children are essentialists, they assume that external characteristics of the group reflect innate, internal characteristics (Bigler & Liben, 2006). For example, although sexualized girls and nonsexualized girls only differ in their clothing and makeup, children may assume that sexualized girls differ in personality traits and abilities from nonsexualized girls (e.g., sexualized girls are less smart and less nice than nonsexualized girls). Thus, we predicted that children would associate external clothing and makeup choices (either sexualized or not) with internal character traits such as intelligence, athleticism, and niceness. These traits were predicted because they map onto the stereotypes endorsed by adults (with sexualized women being low in intelligence, competence, and morality).

Current Studies

As stated, the goal of the present research was to examine whether children endorse a within-gender "sexualized girl stereotype," using the developmental intergroup theory as a framework. In Study 1, we examined whether children associate different traits and attributes with sexualized girls versus nonsexualized girls and whether these group–attribute links are predicted by their cognitive development and media exposure. Study 1 used exaggerated depictions of sexualized girls taken from previous research (Starr & Ferguson, 2012), and Study 2 replicated Study 1 using realistic images of girls that children might see in real life.

In addition to hypothesizing that sexualized girls would be perceived overall as less smart, less athletic, and less nice (as a way of being less moral) than nonsexualized girls, we also predicted that girls would show stronger stereotypic responses than boys would. According to gender schema theory, because of increased self-relevance, girls should have more elaborate and well-developed gender schemas about girls than boys should (Bem, 1981). Further, we hypothesized that children's cognitive development, as assessed through multiple-classification skill, would predict their endorsement of the sexualized girl stereotype, such that children with greater classification skills would endorse a stronger within-gender distinction than children with weaker classification skills. It was also predicted that children's sexualized media exposure would predict their endorsement of the sexualized girl stereotype, such that children with more exposure to sexualized media would endorse the sexualized girl stereotype more strongly than children with little exposure to sexualized media. Beyond these specific theory-based moderators, we predicted that older children would endorse the within-gender stereotype more strongly than younger children, partially because (a) other types of cognitive development (such as understanding individual differences and within-group heterogeneity) increases with age, (b) they have more media exposure over time, and (c) they are approaching puberty and thus sexualized behavior may be more salient.

Study 1

Method

Participants

The study was first approved by the university Institutional Review Board (IRB) to be conducted offsite at five public elementary schools in Lexington, Kentucky through their after-school programs. The director of the after-school program also approved the study. The participants in this study were 208 children (103 girls, 105 boys) recruited from their respective programs. Each program served between 80 and 150 children. Thus, consenting rate at each program ranged from 34% to 48%. Approximately 61% of the children who attended the afterschool programs did not participate in the study due to sporadic attendance (this was exacerbated because the data were collected around winter break), no response to the consent form, or a parent or a child declining participation.

All participants were between the ages of 6 and 11 (Grades 1–5). The mean age was 7.88 ($SD = 1.47$). Of the 208 participants, 74% were European American, 12.5% were African American, 3.8% were Latino, 1.4% were Asian, and 6.3% were multiracial. The gender and ethnic composition of the sample matched the composition of the schools. All of the interviewers were female (seven undergraduate students and one graduate student). All data were collected in the spring of 2013.

Procedure

Parental consent forms were passed out to all parents as they arrived at the after-school program to pick up their children. All parents were told a brief description of the study. If parents signed a consent form (either immediately or after taking the forms home and returning them), their children were then approached about participating. Children who gave their assent were then individually taken into a quiet space in the school to conduct the study. They were told that they could stop the study at any time and that their answers were anonymous and confidential.

Children first performed a series of sorting tasks to measure their classification skill and answered a demographic and sexualized media use questionnaire. Second, using a within-subjects design, all children viewed a total of four images of girls (depicted as paper dolls)—two different nonsexualized girls wearing pants and t-shirts and two sexualized girls wearing short skirts and revealing tops—taken from Starr and Ferguson (2012). A very short vignette about each girl accompanied each image (see Starr & Ferguson, 2012, for images). The four vignettes were written specifically for this study, were counterbalanced across each image, and were kept very brief so as to not contain any information about the traits of the girl. For example, one vignette stated, "This is Olivia. She's in middle school. She lives with her parents and her brother." After listening to the vignette, and while looking at the image of the girl, children were asked questions about the traits of the particular girl. Specifically, the research assistant read the vignette and all questions, and recorded the child's responses, which was repeated for all four dolls. The presentation of images was counterbalanced. Upon completion of the measures, children were debriefed and allowed to pick a small toy from a toy box.

Measures

Classification skill. Classification skill was measured using procedures developed by Jones and Bigler (1996). Children were given a series of pictures of gray bears and elephants and brown bears and elephants. They were asked to sort the pictures into two piles according to one dimension (e.g., by either color or animal type) and then asked to explain their sort. The pile was then shuffled and children were asked to sort the pictures into two new piles, according to a new dimension, and asked to justify their sorting. They received 1 point for each correct sort and 1 point for each correct justification. Children were then presented with a 2 × 2 matrix. Following a demonstration sort, in which the researcher sorts the bears and elephants along both color and animal type dimensions simultaneously, children were asked to create their own 2 × 2 matrix using red and yellow boats and cars and asked to justify their reasoning. Again, children were given 1 point for a correct sort and 1 point for a correct justification. Thus, scores ranged from 0 to 6, with higher scores reflecting greater classification skill.

Stereotypic evaluations of sexualized versus nonsexualized girls. While viewing the sexualized and nonsexualized paper dolls, children were asked about the traits of the girls. Specifically, children were shown a paper doll, then while viewing it, answered the questions. Children were asked, "How popular do you think she is?" "How nice do you think she is?" "How smart do you think she is?" and "How athletic do you think she is?" Each item was rated on a Likert scale, ranging from *not at all* (1) to *very much* (4). The measure was accompanied by a visual scale of a series of cups. The calculation of scores is described in the Results and Discussion section.

Children were also asked several open-ended questions to more deeply explore the content of their stereotypes. In response to the questions, "How popular do you think she is?" "How athletic do you think she is?" and "How smart do you think she is?" children were asked, "Why do you think so?" (open-ended questions were not asked about niceness). The research assistant wrote down their answers verbatim. Open-ended answers were coded first using thematic analysis (Braun & Clarke, 2006) to see which themes emerged without a priori assumptions. The answers were iteratively assessed to find the most common core themes. These differed by trait and are described in detail in the Results and Discussion section. A second round of coding looked specifically for responses that reflected attention to sexualization (if they referred to the girl showing skin, accentuating sexual body parts, or being sexy), appearance focus (if they referred to the girl as being attractive or wearing fashionable clothing), or objectification (if they referred to the girl doing any behavior for someone else's benefit). Responses that strictly used the biography as reasoning (e.g., "Because she goes on the school bus a lot") and responses such as, "I don't know," "Just seems like it," and "Because I think so" were not coded. Two coders, a female undergraduate and graduate student, read through every child's answers and coded their answers. Interrater reliability was adequate, $\kappa = .76$, $p < .001$, 95% CI [.73, .78]. Any inconsistencies were discussed until the coders reached complete agreement.

Sexualized media use. In the demographics questionnaire, children were asked to list their three favorite television shows. Children were asked this question using an open-ended format to better assess which shows were most salient to children. Children's responses were coded using ratings for sexual content taken from www.Commonsensemedia.org. This website rates all media used by children (including television, movies, and video games) on domains such as the amount of sexual content, positive messages, and violence. For this study, we focused on their ratings for sexual content, which the website subdivides into milder sexual content (i.e., "sexy stuff," including flirting and references to sexual interest) and more explicit sexual content (i.e., "sex," including kissing and references to sex). Using their codes, scores could range from 1 to 5, with higher scores indicating greater mild sexualization, or 6 to 10, with higher scores indicating greater explicit sexualization (most shows scored below a 5, indicating mild references to sexuality or sexual interest). Two coders rated every child's answers. Interrater reliability was excellent (99%). Coders independently watched and rated two television shows that were not listed on Commonsensemedia.org. For these two shows, interrater reliability was excellent (100%). Ratings for the three shows were averaged together for one overall score, with a possible range from 1 to 10.

Results and Discussion

Overview of Results and Preliminary Analyses

First, we examined boys' and girls' stereotypes about the traits of sexualized girls and whether these stereotypes were moderated by gender and age. Children's quantitative ratings were examined first in a repeated measures multivariate analysis of variance (MANOVA), followed by the analyses of their qualitative responses. Second, we examined whether classification skill and sexualized media use moderated children's endorsement of the sexualized girl stereotype. For analyses that included age as an independent variable within the MANOVA, we divided children into two age groups, based on a median split ($Mdn = 8$). Thus, children were sorted into a 6- to 8-year-old group ($n = 142$) and a 9- to 11-year-old group ($n = 66$). For other analyses that used regression models (i.e., when examining classification skills), age was treated a continuous variable to preserve power.

Preliminary analyses indicated that classification score was positively correlated with age, such that older children had higher classification scores than younger children, $r = .31$, $p < .05$. Classification score did not differ by gender. The sexualized content of media use was also positively correlated with age, with older children watching more sexualized media than younger children, $r = .39$, $p < .05$. Furthermore, boys ($M = 2.06$, $SD = 1.86$) watched more sexualized media than girls ($M = 1.40$, $SD = 1.40$), $t(131) = -2.30$, $p < .05$. To test for order effects of pictures on trait ratings of sexualized and nonsexualized girls, a 2 (condition: sexualized vs. nonsexualized) × 6 (story order) MANOVA was conducted for the four traits. There were no order effects, $F(40, 995) = 1.10$, $p = .31$.

Comparison Between Sexualized and Nonsexualized Girls

To examine whether children endorsed specific traits for sexualized versus nonsexualized girls, and whether this was moderated by age and gender, children's responses to the same traits for the two sexualized girls were averaged together, and the responses to the same traits for the two nonsexualized girls were averaged together (see Table 1). Each trait (popular, athletic, smart, and nice) was analyzed as part of a 2 (gender: boy vs. girl) × 2 (age median split: 6- to 8-year-olds vs. 9- to 11-year-olds) × 2 (condition: sexualized vs. nonsexualized) repeated measures MANOVA (in which the last variable was the within-subjects variable). As hypothesized, the multivariate main effect for condition was significant, $F(4, 201) = 53.75$, $p < .001$, $\eta^2 = .52$; the multivariate interaction between condition and age was significant, $F(4, 201) = 9.86$, $p < .001$, $\eta^2 = .16$; and the multivariate interaction between condition and gender was significant, $F(4, 201) = 3.78$, $p < .01$, $\eta^2 = .07$. Although not related

Table 1

Study 1: Means and Standard Deviations for Trait Ratings for Sexualized and Nonsexualized Girls

	Boys		Girls	
Traits	6–8 years	9–11 years	6–8 years	9–11 years
Popular				
S	3.05 (.85)	3.22 (.79)	3.40 (.74)	3.48 (.70)
NS	2.89 (.70)	2.63 (.68)	2.84 (.78)	2.76 (.60)
Athletic				
S	2.46 (1.05)	2.05 (1.01)	2.42 (1.03)	1.84 (.97)
NS	2.74 (.93)	2.71 (.81)	2.99 (.76)	3.18 (.50)
Smart				
S	3.39 (.66)	2.60 (.79)	3.28 (.75)	2.74 (.87)
NS	3.55 (.62)	3.43 (.47)	3.63 (.44)	3.60 (.38)
Nice				
S	3.21 (.77)	2.62 (.85)	3.38 (.77)	2.60 (.92)
NS	3.53 (.57)	3.31 (.74)	3.81 (.37)	3.76 (.36)

Note. Means range from 1 to 4, with higher scores indicating greater endorsement of the traits. NS = traits associated with the nonsexualized girl; S = the sexualized girl.

to the hypotheses, and thus not discussed further, the multivariate main effect of gender was significant, $F(4, 201) = 2.67$, $p < .05$, $\eta^2 = .05$, as was the multivariate main effect of age, $F(4, 201) = 7.81$, $p < .01$, $\eta^2 = .14$. Only those results that were significant in the multivariate analysis are discussed as significant at the univariate level. The responses to the open-ended follow-up questions are presented after the univariate results for each relevant trait.

Popular. As hypothesized, there was a significant univariate main effect of condition on ratings of popularity, such that sexualized girls were rated as significantly more popular than nonsexualized girls, $F(1, 204) = 48.19$, $p < .001$, $\eta^2 = .19$. There were several interactions that moderated the main effect. First, there was a marginally significant interaction between condition and gender, $F(1, 204) = 3.40$, $p = .07$, $\eta^2 = .02$. Although both genders perceived the sexualized girls to be more popular than the nonsexualized girls, girls made slightly but not significantly more differentiated ratings than boys did (girls' $\eta^2 = .23$ vs. boys' $\eta^2 = .14$). There was also a significant interaction between condition and age, $F(1, 204) = 4.13$, $p < .05$, $\eta^2 = .02$. Although both age groups perceived the sexualized girls to be more popular than the nonsexualized girls, older children made substantially more differentiated ratings than younger children did (older children's $\eta^2 = .37$ vs. younger children's $\eta^2 = .11$).

Children were asked to report *why* they believed each girl was popular. The most common theme that emerged in children's responses (used by 49% of the sample, $n = 101$) was that the sexualized girl had the types of clothes that a prototypically popular person would wear, namely, "fancy" or "expensive" clothes. In contrast, the nonsexualized girl wore "plain" (i.e., unpopular) clothing. For example, one child stated, pointing to the sexualized girl, "Popular girls would dress really fancy and they would have their hair done and they would wear really nice clothes." Another child stated, in reference to the nonsexualized girl, "Well she doesn't—you know how where you see in the movies, they're all wearing skirts and makeup and stuff and she's wearing a jacket and jeans." These types of responses (about the "fanciness" of the sexualized girls' clothes) reflected the notion that the sexualized girl is appearance focused, and this is predictive of being popular.

The second most common theme that emerged for why children perceived the sexualized girl to be more popular than the nonsexualized girl (used by 21% of the sample, $n = 43$) was that the sexualized girl wears sexualized clothing (i.e., that shows a lot of skin) and that is why she is popular, reflecting a theme of sexualization. This theme more closely reflects the idea that the sexualized clothing specifically is what makes a girl popular. For example, one child stated that the sexualized girl is very popular because "she has a short shirt and a short skirt. And she's wearing heels." Another child stated that the sexualized girl is popular because "she has a belly shirt and her shoulders are showing and her hair is well-kept and the popular people in my grade have really well-kept hair." Thus, children believed that sexualized clothing is the type of clothing that popular people would wear.

Athletic. As hypothesized, there was a significant univariate main effect for condition on ratings of atheticism, such that the nonsexualized girls were rated as significantly more athletic than the sexualized girls, $F(1, 204) = 86.40$, $p < .001$, $\eta^2 = .30$. As predicted, there were also significant interactions that moderated the main effect. First, there was a significant interaction between condition and gender, $F(1, 204) = 9.93$, $p < .01$, $\eta^2 = .05$. Girls made substantially more differentiated ratings than boys did (girls' $\eta^2 = .36$ vs. boys' $\eta^2 = .21$). There was also a significant interaction between condition and age, $F(1, 204) = 14.21$, $p < .001$, $\eta^2 = .07$. Older children made substantially more differentiated ratings than younger children did (older children's $\eta^2 = .51$ vs. younger children's $\eta^2 = .15$).

Again, children were asked to report why they believed the girls were athletic. Children most commonly stated that the sexualized girls' clothes were not compatible with doing athletic things, in contrast to the more "sporty" nonsexualized girls' clothes (used by 42% of the sample, $n = 88$). Many children would state, in reference to the sexualized girl that "those are definitely not sport clothes." Another common theme used by children was that sexualized girls do not want to play sports or be sporty, which was reflected in 19% ($n = 39$) of children's answers. One child, for example, stated that the sexualized girl is not athletic because "she'd probably be like one of those girls to say, 'Oh I got mud on my clothes.'" Another child referred to the sexualized girl as a diva, stating, "Divas don't like to go to gyms and stuff. She hangs out with her friends and go [sic] shopping, like she might work out a little to get her shape in." These responses all reflected the idea that sexualized girls are appearance focused and thus too focused on appearance to care about being athletic.

Smart. As hypothesized, there was a significant univariate main effect for condition on ratings of intelligence, such that nonsexualized girls were rated as significantly smarter than sexualized girls, $F(1, 204) = 86.76$, $p < .001$, $\eta^2 = .30$. There was also a significant interaction between condition and age that moderated the main effect, $F(1, 204) = 25.44$, $p < .001$, $\eta^2 = .11$. As predicted, older children made substantially more differentiated ratings than younger children did (older children's $\eta^2 = .50$ vs. younger children's $\eta^2 = .10$). Gender did not moderate ratings.

In response to why they believed the sexualized girls were not smart, 24% ($n = 50$) of children stated that the sexualized girl is too focused on her clothing or looks to pay attention in school, which again reflects the idea that the sexualized girl is appearance focused. For example, one child stated, in reference to the sexualized girl, that "since she wears that clothes, she wouldn't be paying attention, she would only pay attention to the way she dresses perfect." Another child said, "She like worries on her clothes more." Another 12% ($n = 25$) of children stated that the sexualized girls were simply not as smart as the nonsexualized girls. For example, one child stated that the sexualized girl is not smart because "girls that dress like that aren't very smart." Another 10% ($n = 21$) of children stated that sexualized girls are not as smart as nonsexualized girls because they do not *want* to be smart. One child stated, in reference to the sexualized girl, "some popular girls would act dumb." Overall, children's responses seem to reflect the theme that sexualized girls are so appearance focused that they cannot or will not focus on anything else, including academics. One third of children responded with noncodable answers (such as, "It just looks like it").

Nice. Finally, as hypothesized, there was a significant univariate main effect for condition on ratings of niceness, such that nonsexualized girls were rated as significantly nicer than sexualized girls, $F(1, 204) = 109.25$, $p < .001$, $\eta^2 = .35$. There were significant interactions that moderated the main effect. First, there was a significant interaction between condition and gender, $F(1, 204) = 5.40$, $p < .05$, $\eta^2 = .03$. As with other traits, girls made substantially more differentiated ratings than boys did (girls' $\eta^2 = .40$ vs. boys' $\eta^2 = .28$). The interaction between condition and age was also significant, $F(1, 204) = 19.35$, $p < .001$, $\eta^2 = .09$. Older children made substantially more differentiated ratings than younger children did (older children's $\eta^2 = .51$ vs. younger children's $\eta^2 = .19$). Children were not asked to justify their responses for their nice ratings; thus, there are no qualitative data for the nice responses.

Moderators of Stereotype Endorsement

To test the hypotheses that children with more advanced classification skills and children with more sexualized media use would endorse stronger stereotypes about sexualized girls, two regression models were examined. First, because these hypotheses were focused on an overall stereotype and were not examining individual traits of sexualized girls, an overall "sexualized girl stereotype" score was calculated. The score was created by first calculating the difference between the ratings of sexualized girls and nonsexualized girls for each trait, such that higher positive ratings indicated endorsement in the stereotyped direction (e.g., for the popularity rating, we subtracted the ratings for nonsexualized girls from the ratings of sexualized girls). These difference scores were averaged across all four traits for an overall stereotype score. Scores could thus range from −3 to +3. Scores of 0 indicated no differentiation between the girls, and higher positive scores indicated that children endorsed the stereotype that sexualized girls were more popular, but not as nice, smart, or athletic as nonsexualized girls. Overall, reflecting the pervasiveness of the stereotype, 73.6% of children scored above a 0, indicating that they endorsed the sexualized girl stereotype, while 11.1% of children endorsed no stereotype and 15.4% of children endorsed the counterstereotype.

Classification skill. We first examined whether classification skill predicted the extent to which children endorsed the sexualized girl stereotype, over and above the effects of age. A regression model predicting the endorsement of the stereotype was conducted with age in the first step, classification score in the second step, and the Age × Classification Skill interaction in the third step. Results indicated that the overall model was significant, $F(2, 207) = 21.29$, $R^2 = .17$, $p < .001$. Age predicted endorsement of the sexualized girl stereotype, with older children endorsing the sexualized girl stereotype more than younger children, $\beta = .36$, $p < .001$. Once age was accounted for, however, classification skill only marginally predicted children's endorsement of the sexualized girl stereotype, $\beta = .13$, $p = .07$.

Sexualized media use. We next tested the hypothesis that children who watched media with a higher sexualized content would endorse the sexualized girl stereotype more strongly than children who watched media with lower sexualized content. A regression analysis was conducted with age and gender in the first step, sexualized media rating in the second step, and age and gender interactions in the third step (gender was included because there were initial gender differences in media use). Results indicated that the overall model was significant, $F(5, 132) = 5.65$, $R^2 = .18$, $p < 001$. After controlling for gender and age, the average amount of sexualized content in children's favorite media significantly predicted their endorsement of the sexualized girl stereotype, $\beta = .41$, $p < .05$. This was moderated, however, by the significant interaction between gender and sexualized media use ($\beta = -.34$, $p < .04$). Analysis of simple slopes shows that only girls', but not boys', average sexualized media use predicted their endorsement of the sexualized girl stereotype (girls' $\beta = .29$, $p < .01$; boys' $\beta = .05$, $p = .68$).

Study 2

Overall, results from Study 1 supported the hypotheses that elementary-aged boys and girls perceive sexualized girls as distinct from nonsexualized girls. Specifically, we consider this to be a within-gender stereotype of a sexualized girl: She is popular but not very athletic, smart, or nice relative to nonsexualized girls.

Although Study 1 established that children do endorse a stereotype about sexualized girls, the stimuli for Study 1 were an extreme example of sexualized girls (exaggerated paper doll drawings of sexualized girls). It is possible that children's stereotype for sexualized girls may be less prevalent when presented with more realistic depictions. As described above, the first key component for the development of stereotypes is that children perceive the group attributes to be psychologically and perceptually salient (Bigler & Liben, 2006). Thus, it is important to establish that children notice whether a girl is sexualized or not using ecologically valid stimuli. In Study 2, using a within-subjects design, children were shown photos of a real girl dressed in clothes purchased at a children's clothing store. In other words, children were shown photos of a girl who wore sexualized clothing they could easily see in real life and asked about their perceptions of her traits. To further assess the perceptual and psychological salience of sexualized girls using ecologically valid stimuli, children were also shown two music videos taken from a popular children's television channel that differed in the degree to which the characters were sexualized. The videos are important because they present a large degree of visual information, in addition to the clothing of the sexualized girl. It is important to examine whether children notice the sexualization of the girl (and apply stereotypes to her) when she is embedded within other salient and competing information. To date, no known research has examined children's perceptions of sexualized girls using realistic images.

Study 2 also examined whether children who perceive sexualization to be more salient more strongly endorse the sexualized girl stereotype compared to children who perceive sexualization to be less salient. In other words, according to developmental intergroup theory, for children to categorize and stereotype an individual girl, they must be able to perceive whether she is sexualized or not (i.e., the group attribute must be salient). There are likely, however, individual differences in the degree to which sexualization is salient that may predict the degree to which she is stereotyped.

Finally, we examined how the sexualized girl stereotype was related to children's other gender stereotypes. We focused on two distinct types of gender stereotypes: (a) a traditional gender stereotype about the less powerful role of girls and women relative to boys and men and (b) the belief that girls should be appearance focused. We predicted that the endorsement of traditional gender stereotypes would be unrelated to the endorsement of a sexualized girl stereotype. The sexualized girl stereotype, as established in Study 1, involves sexualized girls being perceived as more popular, but not as smart, athletic, or nice as nonsexualized girls. This stands in contrast to the traditional gender stereotype about girls, in which they are perceived as less powerful than boys, but always nice (Albert & Porter, 1983; Miller et al., 2009). We argue that the sexualized girl stereotype is a within-gender stereotype of girls that is independent of how girls in general are viewed relative to boys. We also examined whether children endorsed the belief that girls in general should be pretty—a component of gender stereotypes that asserts that girls should be appearance focused and valued for their appearance. Although being sexualized is typically considered distinct from being appearance focused (Smolak et al., 2014), children's qualitative answers from Study 1 suggest that children at this age equate the two. For example, they perceived sexualized girls to value their appearance because they are wearing "fancy" clothes. It is likely that children who believe that girls *should* be focused on their appearance would more positively value the sexualized girls than children who believe girls should not be focused on their appearance.

Method

Participants

This study was approved by the IRB through the same process as Study 1. The participants in this study were 155 children (83 girls, 71 boys) recruited from after-school program at two public elementary schools in Lexington, Kentucky. The after-school programs are representative of the larger student body at each school in terms of gender and ethnicity. Each after-school program served between 100 and 150 children. Thus, the consenting rate at each program ranged from 60% to 63%, and approximately 37% to 40% of children who attended the programs did not participate in the study due to sporadic attendance, no response to the consent form, or a parent or child declining participation. All participants were between the ages of 6 and 11 (Grades 1–5). The mean age was 7.80 ($SD = 1.31$). Of the 155 participants, 64.5% were European American, 14.8% were African American, 7.1% were Latino, 5.2% were Asian, and 3.9% were multiracial, matching the ethnic composition of the schools. Consenting was the same as in Study 1.

Procedure

Upon completion of consenting and assenting, children first completed the salience task designed to assess whether sexualization among adults is a salient attribute for children. Participants completed measures assessing their endorsement of traditional gender stereotypes and their beliefs that girls should be appearance focused. Second, using a within-subjects design, children viewed two still photos of a sexualized and nonsexualized girl and two music videos differing in sexualization of the main character. Following each photo and music video, participants were asked about the traits of each girl. Children were then debriefed wherein they were told that the study aimed to understand what children think about girls and were provided the opportunity to ask questions. Afterward, children were given a small toy. All data were collected in the fall of 2013.

Stimuli and Measures

Salience of adult sexualization. Children were asked to sort 12 photographs of women into two piles. The photos were chosen by the authors from online advertisements and consisted of 6 sexualized and 6 nonsexualized women, who also differed in terms of posture (arms above head or arms at waist), hair color (brown or blonde), and ethnicity (European American or African American). Five undergraduate research assistants rated all photos on the level of sexualization and attractiveness, on a scale from *not at all* (1) to *very* (5). The sexualized photos were rated as significantly more sexualized than the nonsexualized photos (averaged ratings, respectively: $Ms = 4.31$ and 1.73, $SDs = .41$ and .43), $t(4) = -11.94$, $p < .001$, $d = 6.14$. The photos did not differ on ratings of attractiveness (averaged ratings, respectively: $Ms = 3.83$ and 3.12, $SDs = .39$ and .69), $t(4) = -2.17$, $p = .10$, $d = 1.27$.

Children were presented with all 12 photos in a random order and told to put the "pictures that looked like they go together" into two piles. Children could have made two piles according to ethnicity, hair color, posture, sexualized clothing, or something idiosyncratic. To assess whether children perceived sexualization to be a salient attribute, research assistants documented which photos were sorted into the two piles (based on numbers written on the back of each photo). Children were given a 1 if at least 75% of the photos of the sexualized women were sorted into one pile or a 0 if they sorted by anything other than sexualization (e.g., ethnicity, posture, hair color). All photos were put away after children completed the sorting task.

Traditional gender stereotypes. A slightly modified version of Galambos's Attitudes Toward Women

Scale for Adolescents was used to assess children's traditional stereotypes about women and girls (Galambos, Peterson, Richards, & Gitelson, 1985). The measure consisted of eight items such as, "Boys are better leaders than girls" and "The father should be in charge of making family decisions." Children responded using a 4-point Likert scale, ranging from *strongly disagree* (1) to *strongly agree* (4), with higher scores indicating more traditional attitudes toward women. All scales were accompanied by a visual aid of frowning and smiling faces. The measure had acceptable psychometric qualities ($\alpha = .71$ for boys, $\alpha = .74$ for girls; $\alpha = .67$ for 6- to 8-year-olds, $\alpha = .75$ for 9- to 11-year-olds).

"Girls as appearance-focused" beliefs. To assess whether children believed girls should be appearance-focused, they were asked three items: "How a girl looks is one of the most important things about her," "I only like girls who are pretty," and "It is important that girls look pretty every day." Children responded using a 4-point Likert scale, ranging from *strongly disagree* (1) to *strongly agree* (4), with higher scores indicating greater endorsement of the belief that girls should be appearance focused. The measure had acceptable psychometric qualities ($\alpha = .75$ for boys, $\alpha = .63$ for girls; $\alpha = .67$ for 6- to 8-year-olds, $\alpha = .71$ for 9- to 11-year-olds).

Stereotypic evaluations of sexualized versus nonsexualized girls. Children were presented with two different photographs of one 9-year-old, European American girl, wearing different clothing. All of the clothing, both sexualized and nonsexualized, was purchased from the same national chain children's clothing store at the local mall. Both outfits were matched for how colorful they were and were both considered "trendy" or fashionable. In the sexualized photo, the girl was wearing a crop-top shirt that revealed her midriff, a short skirt, midheeled shoes, and bracelets. In the nonsexualized girl photo, the girl was wearing skinny blue jeans, a long-sleeved pale green shirt decorated with lace, a sequined sparkly scarf, and feminine dress shoes without a heel. Five undergraduate research assistants rated both photos on level of sexualization and attractiveness, on a scale from *not at all* (1) to *very* (5). The sexualized photo was rated as significantly more sexualized than the nonsexualized photo (averaged ratings, respectively: $Ms = 4.6$ and 1.2, $SDs = .55$ and $.45$), $t(4) = -13.88$, $p < .001$, $d = 6.77$. The photos did not differ on ratings of attractiveness (averaged ratings, respectively: $Ms = 4.0$ and 2.4, $SDs = .71$ and 1.34), $t(4) = 1.97$, $p = .12$, $d = 1.49$.

Before being presented with the first photo, children were told,

> We are going to show you a picture of a girl. Her name is Mindy and she has a twin sister. She lives with her parents and sister. She goes to a school similar to this one. Please look at this picture because we're going to ask you questions about her.

When presented with the second photo, children were told,

> Now we are going to show you a picture of Mandy. She is the twin sister of Mindy. She lives with her parents and her sister. Please look at this picture because we're going to ask you questions about her.

The names and the order of the photos was counterbalanced. Children were then presented the photo of the girl for 10 s. Following each photo, children were asked, "How popular do you think she is?" "How nice do you think she is?" "How smart do you think she is?" and "How athletic do you think she is?" which were the same from Study 1. The items were rated on a Likert scale, ranging from *not at all* (1) to *very much* (4).

Children also viewed two music videos taken from the Disney Channel. The videos are based on two of the most popular shows among children of this age group (based on open-ended responses from Study 1). The sexualized music video (based on characters from the show *Shake It Up*) titled, "Fashion Is My Kryptonite," contains girls wearing tight, short clothing. The focus of the video was girls changing into different sexualized outfits and putting on a fashion show. The lyrics focus on being in style; for example a sample of the lyrics is, "We got style in our veins 'cause fashion's what we breathe." The nonsexualized music video (featuring the main character from the show *Good Luck Charlie*) titled, "We Can Change the World," is focused on helping one another and features girls in skinny jeans and t-shirts. The video also contains images of children and teens holding motivational signs. The lyrics focus on taking steps to help each other; for example, a sample of the lyrics is, "We got the world in our hands, now we're going to start a new day." Both videos ranged from 2 to 3 min long. Before being presented with the music video, children were presented with an image of one of the girls from the video. Children were told to pay attention to the girl in the picture because we would be asking questions about her. Each picture was shown for 10 s before the music video. Five undergraduate research assistants rated the music videos on level of sexualization and attractiveness, on a scale from *not at all* (1) to *very* (5). The sexualized music video was rated as significantly more sexualized than the nonsexualized music video (averaged ratings, respectively: $Ms = 3.8$ and 2.0, $SDs = .45$ and 1.0), $t(4) = 4.81$, $p < .01$, $d = 2.32$. The music videos did not differ on ratings of attractiveness of the key characters (averaged ratings, respectively: $Ms = 4.2$ and 4.2, $SDs = .84$ and .45), $t(4) = .00$, $p = 1.0$, $d = 0$.

After viewing each music video, children were asked about the same traits as before. Children were asked, "How popular do you think she is?" "How nice do you think she is?" "How smart do you think she is?" and "How athletic do you think she is?" The items were rated on a Likert scale, ranging from *not at all* (1) to *very much* (4).

Results and Discussion

Overview of Results and Preliminary Analyses

Results were analyzed similarly to Study 1. First, we examined boys' and girls' stereotypes about the traits of sexualized girls and whether these stereotypes were moderated by gender and age. This was examined for the photos and videos in a repeated measures MANOVA. Second, we examined whether the salience of sexualization, traditional gender stereotypes, and the belief that girls should be appearance focused predicted children's endorsement of the sexualized girl stereotype. As in Study 1, in the MANOVA, children were categorized into a 6- to 8-year-old group ($n = 106$) and a 9- to 11-year-old group ($n = 48$). For the analyses that used regression models (i.e., when examining stereotype endorsement), age was treated a continuous variable.

Correlations were analyzed between the nonsexualized photo and nonsexualized music video ratings for each trait, as were correlations between the sexualized photos and music video for each trait. Analyses found that each paired rating was significantly positively correlated, except for the correlation between the nonsexualized music video and photo for ratings of being smart (which was positively, albeit not significantly, correlated).

Based on which stimuli the child saw first, there were order effects for the sexualized girl stereotype, $F(3, 154) = 3.11$, $p < .05$, $\eta^2 = .06$. Post hoc tests indicated that children who saw the nonsexualized video first endorsed the overall sexualized girl stereotype less strongly than everyone else. To examine whether order effects changed our results, we included stimuli order in the large MANOVA reported below. Including it, however, did not change the results and thus it was omitted in the final analysis to reduce the complexity of the model. To assess whether knowledge of the television show characters from the music videos predicted children's video-based sexualized girl stereotype (calculated only from their difference ratings of the videos), an ANOVA indicated no differences in children's ratings based on knowledge of the characters.

Overall, 63 (41%) children sorted the adult photos by sexualization. Of the 63 children, significantly more girls ($n = 41$, 49%) than boys ($n = 22$, 31%) perceived sexualization to be a salient attribute, $\chi^2(1, n = 153) = 11.78$, $p < .01$, $\phi = .30$. There were no differences in salience of adult sexualization between the younger children ($n = 40$, 38%) and older children ($n = 23$, 48%), $\chi^2(1, n = 153) = .56$, $p = .46$, $\phi = .07$. Boys ($M = 2.56$, $SD = .55$) endorsed stronger traditional gender stereotypes than girls ($M = 2.29$, $SD = .53$), $t(153) = 3.06$, $p < .005$, $d = .50$; there were no differences by age. The belief that girls should be appearance-focused ($M = 2.21$, $SD = .88$) did not differ by gender, yet it was negatively correlated with age, such that younger children had higher scores on the measure than older children ($r = -.16$, $p = .05$).

Comparison Between Sexualized and Nonsexualized Girls

A 2 (gender: boy vs. girl) × 2 (age median split: 6- to 8-year-olds vs. 9- to 11-year-olds) × 2 (condition: sexualized vs. nonsexualized) × 2 (stimuli: music videos vs. photos) repeated measures MANOVA (in which the last two variables were within-subjects variables) was performed for the four traits (see Table 2). Results indicated that the multivariate main effect of condition was significant, $F(4, 144) = 12.49$, $p < .001$, $\eta^2 = .26$. There was a significant multivariate interaction between condition and stimuli, $F(4, 144) = 3.10$, $p < .05$, $\eta^2 = .08$. There was a significant multivariate interaction between condition and age, $F(4, 144) = 2.33$, $p = .05$, $\eta^2 = .06$. Lastly, there was a significant four-way interaction between condition, stimuli, age, and gender, $F(4, 144) = 2.59$, $p < .05$, $\eta^2 = .07$. Although not part of the hypotheses, and thus not discussed further, there was also a significant multivariate main effect of stimuli, $F(4, 144) = 35.72$, $p < .001$, $\eta^2 = .50$, such that children rated the girls

Table 2

Study 2: Means and Standard Deviations for Trait Ratings for Sexualized and Nonsexualized Girls

	Photographs				Videos			
	Boys		Girls		Boys		Girls	
	6–8 years	9–11 years	6–8 years	9–11 years	6–8 years	9–11 years	6–8 years	9–11 years
Popular								
S	1.98 (.99)	1.76 (1.04)	2.k04 (.98)	1.96 (1.04)	2.59 (.64)	2.55 (.74)	2.74 (.64)	2.77 (.43)
NS	1.88 (.95)	1.41 (.91)	1.80 (.98)	1.92 (.80)	2.57 (.74)	2.73 (.55)	2.72 (.45)	2.73 (.53)
Athletic								
S	1.80 (1.19)	1.48 (.81)	1.46 (1.07)	1.74 (1.11)	1.76 (1.16)	1.50 (.91)	1.84 (1.12)	1.54 (1.07)
NS	1.73 (1.06)	1.82 (.96)	1.89 (1.04)	1.96 (.96)	1.85 (1.07)	1.73 (.99)	1.95 (1.09)	1.88 (.91)
Smart								
S	2.33 (.90)	2.19 (.87)	2.56 (.71)	2.15 (.97)	2.53 (.74)	2.14 (.77)	2.53 (.78)	2.50 (.65)
NS	2.53 (.71)	2.32 (.89)	2.59 (.68)	2.65 (.56)	2.73 (.64)	2.82 (.50)	2.70 (.63)	2.77 (.43)
Nice								
S	2.33 (.83)	1.95 (.97)	2.56 (.66)	2.23 (.99)	2.69 (.55)	2.32 (.78)	2.74 (.55)	2.77 (.51)
NS	2.59 (.64)	2.41 (.80)	2.66 (.67)	2.77 (.43)	2.86 (.41)	2.64 (.58)	2.91 (.34)	2.73 (.45)

Note. Means range from 1 to 4, with higher scores indicating greater endorsement of the traits. NS = traits associated with the nonsexualized girl; S = the sexualized girl.

in the videos as more popular, nicer, and smarter than the girl in the photo (likely a function of greater familiarity). Only those results that were significant in the multivariate analysis are discussed as significant at the univariate level.

There was a main effect of condition on ratings of being athletic $F(1, 147) = 10.13$, $p < .01$, $\eta^2 = .06$; being smart, $F(1, 147) = 38.88$, $p < .001$, $\eta^2 = .18$; and being nice, $F(1, 147) = 26.19$, $p < .001$, $\eta^2 = .15$. As predicted, these findings replicate those of Study 1; there was no effect, however, of condition on ratings of popularity. There were also several interactions that moderated and qualified these main effects. Specifically, the Condition × Age interaction was significant for ratings of being athletic, $F(1, 147) = 5.39$, $p < .05$, $\eta^2 = .04$. Although both age groups of children rated the sexualized girls (in both the photos and videos) as less athletic than the nonsexualized girls, only older children significantly differentiated between the girls (older children's $\eta^2 = .21$ vs. younger children's $\eta^2 = .01$).

There was also a significant interaction between condition and age for being smart, $F(1, 147) = 6.05$, $p < .05$, $\eta^2 = .04$, with older children again differentiating more than younger children (older children's $\eta^2 = .34$ vs. younger children's $\eta^2 = .08$). This effect is qualified, however, by the significant four-way interaction between condition, stimuli, gender, and age, $F(1, 147) = 6.44$, $p < .05$, $\eta^2 = .04$. Tests of simple effects indicated that although all children responded in the same direction, older boys were differentiating between the girls in the videos ($p < .001$, $\eta^2 = .41$), whereas older girls were differentiating between the girls in the photos ($p < .01$, $\eta^2 = .26$).

Finally, there was a significant interaction between condition and stimuli for ratings of being nice, $F(1, 147) = 3.89$, $p = .05$, $\eta^2 = .03$. Tests of simple effects indicated that children differentiated more between the girls when looking at photos than the music videos (pictures $\eta^2 = .12$ vs. videos $\eta^2 = .06$).

Salience of Adult Sexualization

To test the hypothesis that children who perceived sexualization of adults to be a salient attribute would more strongly endorse the sexualized girl stereotype, two 2 (salience of sexualization: yes vs. no) × 2 (gender: boy vs. girl) × 2 (age: 6- to 8-year-olds vs. 9- to 11-year-olds) ANOVAs were conducted. One sexualized girl stereotype score was calculated for the photo stimuli and one score was calculated for the music video stimuli (each calculated the same as in Study 1). These scores were kept separate in this analysis (instead of collapsing into one overall score) because the stimuli differed in the amount of competing perceptual information.

Results indicated that salience was not related to the photograph-based sexualized girl stereotype, $F(1, 120) = .002$, $p = .96$, $\eta^2 = .00$. There was, however, a main effect of salience on the video-based sexualized girl stereotype, $F(1, 121) = 4.56$, $p < .05$, $\eta^2 = .04$. Specifically, children that sorted based on sexualization (i.e., for whom sexualization was a salient attribute) endorsed the sexualized girl stereotype in response to the music videos more than children that did not sort based on sexualization (*M*s = .20 and .06, *SD*s = .38 and .37, respectively).

Relationship Between Sexualized Girl Ratings and Gender Stereotypes and Beliefs

Traditional gender stereotypes. To test the hypothesis that the endorsement of traditional gender stereotypes would be distinct from children's sexualized girl stereotypes, we examined the overall sexualized girl stereotype score (we collapsed across the photos and videos, because the ratings were significantly correlated with one another and there was no theoretical reason to keep separate). A regression analysis was conducted in which age and gender were entered in the first step, and the traditional gender stereotype score was entered in the second step, and the Traditional Gender Stereotype Score × Gender interaction was entered in the third step predicting the sexualized girl stereotype. In the final model, $F(4, 153) = 4.03$, $R^2 = .10$, $p < .005$, age significantly predicted the sexualized girl stereotype ($\beta = .32$, $p < .01$). Older children more strongly endorsed the stereotype than younger children. As hypothesized, however, traditional gender stereotypes did not predict the sexualized girl stereotype ($\beta = .35$, $p = .18$), nor did the interaction between traditional gender stereotypes and gender ($\beta = -.27$, $p = .30$).

"Girls as appearance focused" beliefs. To test the hypothesis that children who believed that girls should be appearance focused would hold positive attitudes toward the sexualized girls, we were interested in just children's ratings of the sexualized girls (and not their ratings of the nonsexualized girls). Therefore, we calculated ratings for each of the four traits of the sexualized girls, averaging across the sexualized photo and music video. Each trait rating for the sexualized girl in the photo and music video was significantly positively correlated (popular: $r = .22$, $p < .01$; athletic: $r = .54$, $p < .001$; nice: $r = .34$, $p < .001$; smart: $r = .36$, $p < .001$). We analyzed each sexualized girl trait in a regression analysis, in which we controlled for age and gender in the first step, and entered the "girls as appearance focused" belief score in the second step. Results indicated that, controlling for gender and age, children who more strongly endorsed the belief that girls *should* be appearance focused rated the sexualized girls as more popular ($\beta = .21$, $p < .01$), model $F(3, 152) = 2.81$, $R^2 = .05$, $p < .05$; more athletic ($\beta = .19$, $p < .05$), model $F(3, 152) = 2.99$, $R^2 = .06$, $p < .05$; smarter ($\beta = .29$, $p < .01$), model $F(3, 152) = 7.83$, $R^2 = .14$, $p < .001$; and nicer ($\beta = .19$, $p < .05$), model $F(3, 152) = 8.25$, $R^2 = .14$, $p < .01$, than children who did not endorse those beliefs.

General Discussion

The purpose of the current studies was to examine whether children have developed a within-gender stereotype of girls, distinct from traditional gender stereotypes, referred to here as the sexualized girl stereotype. Across two studies, we found evidence that children do hold unique stereotypes about sexualized girls. Specifically, sexualized girls are decidedly less athletic, nice, and intelligent than nonsexualized girls. Children apply this stereotype to exaggerated depictions of sexualized girls (Study 1), realistic photos of girls in easily attainable clothes (Study 2), and girls in music videos (Study 2). When the sexualized girl is an exaggerated depiction, she is also perceived as more popular than the nonsexualized girl.

Furthermore, based on the qualitative data, children in middle childhood seem to assume that sexualized girls are appearance focused (or that they are choosing appearance concerns over concerns about being intelligent and athletic). For example, statements such as, "She'd probably be like one of those girls to say, 'Oh I got mud on my clothes,'" "Since she wears that clothes, she wouldn't be paying attention, she would only pay attention to the way she dresses perfect," "She like worries on her clothes more," and "Popular girls would dress really fancy and they would have their hair done and they would wear really nice clothes," reflect a focus on appearance concerns. Indeed, for children who have not yet entered puberty, being appearance focused is a much more prevalent characteristic of sexualized girls than being sexual per se. Interestingly, these constructs are not synonymous and are theoretically orthogonal (Smolak et al., 2014), but for children who are not yet sexual, being appearance focused is at least a component of being sexualized. Sexualized girls, therefore, in some ways embody the Western feminine norm of focusing on their appearance (Mahalik et al., 2005). They defy other feminine norms, however, by being less nice than other girls (Mahalik et al., 2005).

Based on developmental intergroup theory (Bigler & Liben, 2006), because children made distinctions between the two types of girls across various stimuli, being sexualized or not was obviously a perceptually and psychologically salient attribute for elementary school children. Even when their clothing was not pointed out or mentioned, children encoded the differences and made trait attributions based on those differences. Furthermore, as predicted by developmental intergroup theory, the girls' short skirts and cropped tops were not simply external clothing choices, but represented internal character traits such as intelligence, athleticism, and niceness.

Although the majority of children (more than 75%) endorsed this sexualized girl stereotype, there were individual differences in the strength of this endorsement. First, there were individual differences in how salient sexualization was as a characteristic. Children who perceived sexualization to be a salient characteristic more strongly endorsed the sexualized girl stereotype when presented with distracting, competing information (e.g., in a busy music video). When looking at still images with fewer distractions, almost all children endorsed the stereotype, regardless of how salient sexualization was to them personally. When there is other information that can be used as a basis of judgment, as in the video, it was only the children who sorted by sexualization that differentiated between the sexualized girl and nonsexualized girl.

Second, as predicted, the sexualized girl stereotype was, at times, more strongly endorsed by girls than boys. According to gender schema theory, it is more important for girls than boys to pay attention to information about girls as a way to create a self-relevant gender schema (e.g., Bem, 1981; Martin & Halverson, 1981). Thus, seeing sexualized girls may activate a broader gender schema about girls, which leads girls, in some cases, to have stronger stereotypic responses to sexualized girls than boys do.

Relatedly, although we predicted that children who watched more sexualized media would be more likely to endorse the sexualized girl stereotype than children who watched less sexualized content, this effect was only true for girls. The more sexualized media that girls consumed, the greater their distinctions were between the two types of girls. This suggests that girls are especially attuned, relative to boys, to environmental messages about sexualized girls versus nonsexualized girls. This is also supported by the finding in which sexualization was a more salient attribute for girls than for boys (from the sorting task in Study 2). The implication of this finding is that girls seem particularly vulnerable to media messages about sexualized girls, likely because of the greater self-relevance of the messages. Previous research has pointed out that repeated sexualization messages are associated with negative outcomes for girls, such as increases in body shame, surveillance, and disordered eating, as well as decreases in academic performance, body esteem, and psychological well-being (Fredrickson et al., 1998; Lindberg, Grabe, & Hyde, 2007 McKenney & Bigler, 2014a, 2014b; McKinley, 1999; Muehlenkamp, Swanson, & Brausch, 2005; Tiggemann & Lynch, 2001). The current studies extend this previous research by indicating that even when boys and girls are assessed on a comparable outcome, stereotypes about girls, girls are the most vulnerable to sexualization messages.

The sexualized gender stereotype was also consistently more strongly endorsed by older children than younger children. We presume that there are social, biological, and cognitive reasons for this effect. First, older children have likely been exposed to more sexualized messages than younger children, which may have increased the number of times they have seen the group–attribute links regarding sexualized girls (Bigler & Liben, 2006). This assumption is supported by the general finding in the current studies linking exposure to sexualized media with the endorsement of the sexualized girl stereotype (Study 1). It also supports previous work showing that girls who see more sexualized girls and women in the media show more negative psychological outcomes (e.g., lower self-esteem and greater depression; American Psychological Association, 2007).

Second, research has shown that as children reach adolescence and approach puberty, they become more aware of and concerned about their body image (Lindberg et al., 2007). They may also begin to be more concerned with how they appear to the other gender (at least for heterosexual youth), including how they can best attract the interest of potential suitors. Thus, the greater endorsement of the sexualized girl stereotype among older children compared to younger children may reflect impending pubertal (i.e., sexual) interests. To examine this, future research should explore whether children who have already reached puberty endorse the sexualized girl stereotype more strongly than their age-mates who have not yet reached puberty. Future research should also examine whether post-pubertal youth associate sexualized girls with more sexual behaviors, rather than just appearance concerns (which was the primary focus for prepubertal children in middle childhood).

Third, although the effect was only marginally significant, children who had greater classification abilities were slightly more likely to endorse the sexualized girl stereotype, over and above the effects of age, than children with less developed abilities (Study 1). This supports the hypothesis that children's cognitive development (to a degree) facilitates their categorization of sexualized girls into a distinct category of girls. Future research should explore the role of cognitive development in influencing stereotype endorsement. It may be that multiple-classification ability is not the most predictive cognitive skill associated with within-gender stereotypes. Or, it may be that multiple-classification skill, in combination with individual difference variables such as gender schematicity (or the degree to which gender is chronically accessible to the child; Levy, 1994), best predicts children's stereotypes.

Regardless of the exact reasons for this developmental trend, the implications for development across middle childhood are important. By the end of elementary school, children have distinct stereotypes about what a sexualized girl is like. Importantly, these negative stereotypes extend to a girl wearing clothes that are easily available from an extremely popular store in the local mall, in other words, a girl who could easily be in their class at school. Because these stereotypes are associated with academic abilities, sports participation, and prosocial development, they may be particularly harmful for elementary-aged children who are in the process of developing these abilities.

Ironically, although most of the stereotypes about sexualized girls are negative, there is variability in how positively they are perceived. Both boys and girls who believe that girls should always be pretty and valued for their appearance are more likely to perceive sexualized girls positively overall. The exact impact of these attitudes remains unclear. Future research should explore the long-term impacts of valuing the appearance of girls and the associated belief that being sexualized is positive for girls. For girls, these attitudes are most likely associated with greater body shame and surveillance (Fredrickson et al., 1998); for boys, they may be associated with greater acceptance of sexual harassment toward girls (Brown & L'Engle, 2009). Future longitudinal work should assess the impact of endorsing a sexualized girl stereotype in elementary school.

Future research should also examine how girls internalize these positive and negative stereotypes. For example, if girls assume that sexualized girls are supposed to "act dumb" (as stated by one participant) and girls want to look like the sexualized girl because of her popularity, girls may downplay their own intelligence and underperform at school (McKenney & Bigler, 2014a, 2014b). This is particularly concerning considering the gender gap that still exists in science, technology, engineering, and mathematics fields that begins to emerge in early adolescence (Robnett & Leaper, 2013). In addition, girls may begin to drop out of athletics as they identify more and more with the sexualized girl (who is not supposed to be athletic). This is problematic because girls may then forego the many positive outcomes associated with girls' involvement in sports, such as positive self-esteem and body esteem (Kane et al., 2007). Future research should also examine how other children and adults view girls who internalize and embody these stereotypes. For example, if another student perceives a girl to be sexualized, he or she may behave toward that girl in accordance with the sexualized girl stereotype and thus treat her as less intelligent, athletic, and nice. Research on gender stereotypes has consistently shown that children often act in accordance with the stereotypes (e.g., Leaper & Brown, 2014), further contributing to the existing gender gaps in academics and athletics. Future research should examine individual characteristics that could lead to the endorsement and internalization of the sexualized girl stereotype, such as parental socialization, previous academic success, and need to belong. The implications for boys are just as important. It is important to examine how boys' behaviors and beliefs may be influenced by their attitudes about girls. This is likely especially influential as children approach puberty and begin to reengage in cross-sex interactions.

Of course, this study is not without limitations. First, each trait was assessed by only one item. Furthermore, the stimuli in the study may not have been equally gender-typed. For example, the nonsexualized girl in the photo and paper doll was wearing pants, which may be seen as less feminine than a skirt. Future research should examine whether the sexualized girl stereotype is applied to girls that are wearing dresses or other types of clothing that are still highly feminine but not sexualized. This study was also conducted with a largely European American, middle-class sample using European American girls as stimuli. Future research should examine whether this stereotype is endorsed by a more diverse group of girls, and whether it is applied to a more diverse group of targets. Furthermore, future research needs to examine other possible mediators (such as concern for popularity) between children's endorsement of the sexualized girl stereotype and how children are impacted by these stereotypes. Despite these limitations, however, this study suggests that children do indeed have elaborate and detailed stereotypes about sexualized girls. Considering the increasingly common culture of sexualized girls in the media, it is critical that researchers more fully address the impact of these messages on the developing child.

References

Aboud, F. E. (1989). *Children and prejudice*. Oxford, UK: Blackwell.

Albert, A. A., & Porter, J. R. (1983). Age patterns in the development of children's gender-role stereotypes. *Sex Roles*, *9*, 59–67. doi:10.1007/BF00303110

American Psychological Association. (2007). *Report of the APA task force on the sexualization of girls*. Washington, DC: Author. Retrieved from http://www.apa.org/pi/women/programs/girls/report.aspx

Aubrey, J., & Harrison, K. (2004). The gender-role content of children's favorite television programs and its links to their gender-related perceptions. *Media Psychology*, *6*, 111–146. doi:10.1207/s1532785xmep0602_1

Aubrey, J. S., Hopper, K. M., & Mbure, W. G. (2011). Check that body! The effects of sexually objectifying music videos on college men's sexual beliefs. *Journal of Broadcasting & Electronic Media*, *55*, 360–379. doi:10.1080/08838151.2011.597469

Bem, S. L. (1981). Gender schema theory: A cognitive account of sex typing. *Psychological Review*, *88*, 354–364. doi:10.1037/0033-295X.88.4.354

Bigler, R. S. (1995). The role of classification skill in moderating environmental influences on children's gender stereotyping: A study of the functional use of gender in the classroom. *Child Development*, *66*, 1072–1087. doi:10.2307/1131799

Bigler, R. S., Arthur, A. E., Hughes, J. M., & Patterson, M. M. (2008). The politics of race and gender: Children's perceptions of discrimination and the US presidency. *Analyses of Social Issues and Public Policy*, *8*, 83–112. doi:10.1111/j.1530-2415.2008.00161.x

Bigler, R. S., & Liben, L. S. (1992). Cognitive mechanisms in children's gender stereotyping: Theoretical and educational implications of a cognitive-based intervention. *Child Development*, *63*, 1351–1363. doi:10.2307/1131561

Bigler, R. S., & Liben, L. S. (2006). A developmental intergroup theory of social stereotypes and prejudice. *Advances in Child Development and Behavior*, *34*, 39–89. doi:10.1016/S0065-2407(06)80004-2

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*, 77–101. doi:10.1191/1478088706qp063oa

Brown, J. D., & L'Engle, K. L. (2009). X-rated: Sexual attitudes and behaviors associated with U.S. early adolescents' exposure to sexually explicit media. *Communication Research*, *36*, 129–151. doi:10.1177/0093650208326465

Browne, B. A. (1998). Gender stereotypes in advertising on children's television in the 1990s: A cross-national analysis. *Journal of Advertising*, *27*, 83–96. doi:10.1080/00913367.1998.10673544

Conrad, K., Dixon, T., & Zhang, Y. (2009). Controversial rap themes, gender portrayals and skin tone distortion: A content analysis of rap music videos. *Journal of Broadcasting & Electronic Media*, *53*, 134–156. doi:10.1080/08838150802643795

Daniels, E. A. (2009). Sex objects, athletes, and sexy athletes: How media representations of women athletes can impact adolescent girls and college women. *Journal of Adolescent Research*, *24*, 399–422. doi:10.1177/0743558409336748

Daniels, E. A., & Zurbriggen, E. L. (2014). The price of sexy: Viewers' perceptions of a sexualized versus nonsexualized Facebook profile photograph. *Psychology of Popular Media Culture*. Advance online publication. doi:10.1037/ppm0000048

Downs, E., & Smith, S. L. (2010). Keeping abreast of hypersexuality: A video game character content analysis. *Sex Roles*, *62*, 721–723. doi:10.1007/s11199-009-9637-1

Dunn, J., Lewis, V., & Patrick, S. (2010). The idealization of thin figures and appearance concerns in middle school children. *Journal of Applied Biobehavioral Research*, *15*, 134–143. doi:10.1111/j.1751-9861.2010.00057.x

Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin*, *136*, 103–127. doi:10.1037/a0018053

Fredrickson, B. L., & Roberts, T. (1997). Objectification theory: Toward understanding women's lived experiences and mental health risks. *Psychology of Women Quarterly*, *21*, 173–206. doi:10.1111/j.1471-6402.1997.tb00108.x

Fredrickson, B. L., Roberts, T., Noll, S. M., Quinn, D. M., & Twenge, J. M. (1998). That swimsuit becomes you: Sex differences in self-objectification, restrained eating, and math performance. *Journal of Personality and Social Psychology*, *75*, 269–284. doi:10.1037/0022-3514.75.1.269

Galambos, N. L., Peterson, A. C., Richards, M., & Gitelson, I. B. (1985). The Attitudes Toward Women Scale for Adolescents (AWSA): A study of reliability and validity. *Sex Roles*, *13*, 343–356. doi:10.1007/BF00288090

Glick, P., Larsen, S., Johnson, C., & Branstiter, H. (2005). Evaluations of sexy women in low- and high-status jobs. *Psychology of Women Quarterly*, *29*, 389–395. doi:10.1111/j.1471-6402.2005.00238.x

Golombok, S., Rust, J., Zervoulis, K., Croudace, T., Golding, J., & Hines, M. (2008). Developmental trajectories of sex-typed behavior in boys and girls: A longitudinal general population study of children aged 2.5–8 years. *Child Development*, *75*, 1583–1593. doi:10.1111/j.1467-8624.2008.01207.x

Goodin, S. M., Denburg, A. V., Murnen, S. K., & Smolak, L. (2011). "Putting on" sexiness: A content analysis of the presence of sexualizing characteristics in girls' clothing. *Sex Roles, 65,* 1–12. doi:10.1007/s11199-011-9966-8

Graff, K., Murnen, S. K., & Smolak, L. (2012). Too sexualized to be taken seriously? Perceptions of a girl in childlike vs. sexualizing clothing. *Sex Roles, 66,* 764–775. doi:10.1007/s11199-012-0145-3

Grauerholz, E., & King, A. (1997). Prime time sexual harassment. *Violence Against Women, 3,* 129–148. doi:10.1177/1077801297003002003

Hamilton, M. C., Anderson, D., Broaddus, M., & Young, K. (2006). Gender stereotyping and under-representation of female characters in 200 popular children's picture books: A twenty-first century update. *Sex Roles, 55,* 757–765. doi:10.1007/s11199-006-9128-6

Hatton, E., & Trautner, M. N. (2011). Equal opportunity objectification? The sexualization of men and women on the cover of *Rolling Stone. Sexuality & Culture, 15,* 256–278. doi:10.1007/s12119-011-9093-2

Herbozo, S., Tantleff-Dunn, S., Gokee-Larose, J., & Thompson, J. K. (2004). Beauty and thinness messages in children's media: A content analysis. *Eating Disorders, 12,* 21–34. doi:10.1080/10640260490267742

Jones, D. C., & Crawford, J. K. (2006). The peer appearance culture during adolescence: Gender and body mass variations. *Journal of Youth and Adolescence, 35,* 257–269. doi:10.1007/s10964-005-9006-5

Jones, L. C., & Bigler, R. S.. (1996). *Cognitive-developmental mechanisms in the revision of gender stereotypic beliefs.* Poster presented at the biennial meeting of the conference on Human Development, Birmingham, AL.

Kane, M. J., LaVoi, N., Wiese-Bjornstal, D., Duncan, M., Nichols, J., Pettee, K., & Ainsworth, B.. (2007). *The 2007 Tucker Center Research on Girls & Women in Sport.* University of Minnesota Digital Conservancy. Retrieved from http://purl.umn.edu/48566

Kuhn, D., Nash, S. C., & Brucken, L. (1978). Sex role concepts of two- and three-year olds. *Child Development, 49,* 445–451. doi:10.2307/1128709

Lacroix, C. (2004). Images of animated others: The orientalization of Disney's cartoon heroines from the Little Mermaid to the Hunchback of Notre Dame. *Popular Communication, 2,* 213–229. doi:10.1207/s15405710pc0204_2

Leaper, C., & Brown, C. S. (2014). Sexism in school contexts. In L. S. Liben & R. S. Bigler (Eds.), *The role of gender in educational contexts and outcomes: Advances in child development and behavior.* London, UK: Elsevier.

Lerum, K., & Dworkin, S. L. (2009). "Bad girls rule": An interdisciplinary feminist commentary on the report of the APA task force on the sexualization of girls. *Journal of Sex Research, 46,* 250–263. doi:10.1080/00224490903079542

Levy, G. D. (1994). High and low gender schematic children's release from proactive interference. *Sex Roles, 30,* 93–108. doi:10.1007/BF01420742

Liben, L. S., Bigler, R. S., & Krogh, H. R. (2001). Pink and blue collar jobs: Children's judgments of job status and job aspirations in relation to sex of worker. *Journal of Experimental Child Psychology, 79,* 346–363. doi:10.1006/jecp.2000.2611

Lindberg, S. M., Grabe, S., & Hyde, J. S. (2007). Gender, pubertal development, and sexual harassment predict objectified body consciousness in early adolescence. *Journal of Research on Adolescence, 17,* 723–742. doi:10.1111/j.1471-6402.2007.00350.x

Mahalik, J. R., Morray, E. B., Coonerty-Femiano, A., Ludlow, L. H., Slattery, S. M., & Smiler, A. (2005). Development of the Conformity to Feminine Norms Inventory. *Sex Roles, 52,* 417–435. doi:10.1007/s11199-005-3709-7

Martin, C. L., & Halverson, C. F. (1981). A schematic processing model of sex typing and stereotyping in children. *Child Development, 52,* 1119–1134. doi:10.2307/1129498

McKenney, S. J., & Bigler, R. S. (2014a). High heels, low grades: Internalized sexualization and academic orientations among adolescent girls. *Journal of Research on Adolescence.* Advance online publication. doi:10.1111/jora.12179

McKenney, S., & Bigler, R. S. (2014b). Internalized sexualization and its relation to sexualized appearance, body surveillance, and body shame among early adolescent girls. *Journal of Early Adolescence.* Advance online publication. doi:10.1177/0272431614556889

McKinley, N. M. (1999). Women and objectified body consciousness: Mothers' and daughters' body experience in cultural, developmental, and familial context. *Developmental Psychology, 35,* 760–769. doi:10.1037/0012-1649.35.3.760

Miller, C., Lurye, L. E., Zosuls, K. M., & Ruble, D. N. (2009). Accessibility of gender stereotype domains: Developmental and gender differences in children. *Sex Roles, 60,* 870–881. doi:10.1007/s11199-009-9584-x

Muehlenkamp, J. J., Swanson, J. D., & Brausch, A. M. (2005). Self-objectification, risk taking, and self-harm in college women. *Psychology of Women Quarterly, 29,* 24–32. doi:10.1111/j.1471-6402.2005.00164.x

Near, C. E. (2013). Selling gender: Associations of box art representation of female characters with sales for teen- and mature-rated video games. *Sex Roles, 68,* 252–269. doi:10.1007/s11199-012-0231-6

Piaget, J., & Inhelder, B. (1969). *The psychology of the child.* New York, NY: Basic Books.

Robnett, R. D., & Leaper, C. (2013). Friendship groups, personal motivation, and gender in relation to high school students' STEM career interest. *Journal of Research on Adolescence, 23,* 652–664. doi:10.1111/jora.12013

Rudman, L. A., & Borgida, E. (1995). The afterglow of construct accessibility: The behavioral consequences of priming men to view women as sexual objects. *Journal of Experimental Social Psychology, 31,* 493–517. doi:10.1006/jesp.1995.1022

Smolak, L., Murnen, S. K., & Myers, T. A. (2014). Sexualizing the self: What college women and men think about and do to be "sexy." *Psychology of Women Quarterly, 38,* 379–397. doi:10.1177/0361684314524168

Starr, C. R., & Ferguson, G. M. (2012). Sexy dolls, sexy grade-schoolers? Media & maternal influences on young girls' self-sexualization. *Sex Roles*, *67*, 463–476. doi:10.1007/s11199-012-0183-x

Tibbetts, S. L. (1975). Sex-role stereotyping in the lower grades: Part of the solution. *Journal of Vocational Behavior*, *6*, 255–261. doi:10.1016/0001-8791(75)90051-2

Tiggemann, M., & Lynch, J. E. (2001). Body image across the life span in adult women: The role of self-objectification. *Developmental Psychology*, *37*, 243–253. doi:10.1037/0012-1649.37.2.243

Tiggemann, M., & Slater, A. (2014). Contemporary girlhood: Maternal reports on sexualized behaviour and appearance concern in 4–10 year-old girls. *Body Image*, *11*, 396–403. doi:10.1016/j.bodyim.2014.06.007

Turrill, D.. (2013). *Free to move between screens: The cross-platform report.* Neilsenwire. Retrieved from http://www.nielsen.com/content/dam/corporate/uk/en/documents/Nielsen-March-2013-Cross-Platform-Report.pdf

Ward, L. M. (1995). Talking about sex: Common themes about sexuality in the prime-time television programs children and adolescents view most. *Journal of Youth and Adolescence*, *24*, 595–615. doi:10.1007/BF01537058

Ward, L. M. (2002). Does television exposure affect emerging adults' attitudes assumptions about sexual relationships? Correlational and experimental confirmation. *Journal of Youth and Adolescence*, *31*, 1–15. doi:10.1023/A:1014068031532

Williams, J. E., Bennett, S. M., & Best, D. L. (1975). Awareness and expression of sex stereotypes in young children. *Developmental Psychology*, *11*, 635–642. doi:10.1037/0012-1649.11.5.635

Child Development, September/October 2015, Volume 86, Number 5, Pages 1623–1638

Distal Communication by Chimpanzees (*Pan troglodytes*): Evidence for Common Ground?

David A. Leavens
University of Sussex

Lisa A. Reamer and Mary Catherine Mareno
Michale E. Keeling Center for Comparative Medicine and Research

Jamie L. Russell and Daniel Wilson
Georgia State University

Steven J. Schapiro
Michale E. Keeling Center for Comparative Medicine and Research

William D. Hopkins
Georgia State University and Yerkes National Primate Research Center

van der Goot et al. (2014) proposed that distal, deictic communication indexed the appreciation of the psychological state of a common ground between a signaler and a receiver. In their study, great apes did not signal distally, which they construed as evidence for the human uniqueness of a sense of common ground. This study exposed 166 chimpanzees to food and an experimenter, at an angular displacement, to ask, "Do chimpanzees display distal communication?" Apes were categorized as (a) proximal or (b) distal signalers on each of four trials. The number of chimpanzees who communicated proximally did not statistically differ from the number who signaled distally. Therefore, contrary to the claim by van der Goot et al., apes do communicate distally.

Immense theoretical weight has been given to reports of species differences in cognitive abilities between human beings and representatives of our nearest living relatives, the great apes (e.g., Burkart, Hrdy, & van Schaik, 2009; Herrmann, Call, Hernandez-Lloreda, Hare, & Tomasello, 2007; Liszkowski, Schäfer, Carpenter, & Tomasello, 2009; Povinelli, Bierschwale, & Čech, 1999; Tomasello, 2008; Tomasello, Call, & Gluckman, 1997). In particular, it is widely believed that human children, as young as 12 months of age, have a unique capacity for cooperative behavior and that this putative species-unique cooperative motivation forms a developmental platform for more advanced sociocognitive representations (e.g., Tomasello, 2008). However, to our knowledge, none of the many empirical studies purporting to demonstrate an early-developing and human-unique capacity for understanding social signals actually supports the conclusion that humans, alone, have these capacities.

For example, no researcher who has claimed to report a species difference in sociocognitive ability between humans and apes has ever matched the humans and apes on their age or life history stage (Bard & Leavens, 2014). Moreover, no claim of a species difference in social cognition between great apes and humans has ever matched the humans and the apes on their task-relevant pre-experimental

---

This research was supported by NIH grants MH-92923, NS-42867, NS-73134 and HD-60563, and U42-OD-011197. David Leavens was also supported, in part, by an internal grant from the School of Psychology Research Committee, School of Psychology, University of Sussex. The animal care practices and facilities at the University of Texas MD Anderson Cancer Center and the Yerkes National Primate Research Center are fully accredited by the Association for the Assessment and Accreditation of Laboratory Animal Care-International. American Psychological Association and Institute of Medicine guidelines for the treatment of chimpanzees in research were followed during all aspects of this study. We thank four anonymous reviewers for their thoughtful analyses and recommendations on previous versions of this paper.

Correspondence concerning this article should be addressed to David Leavens, School of Psychology, University of Sussex, Falmer, East Sussex, BN1 9QH, United Kingdom. Electronic mail may sent to davidl@sussex.ac.uk.

© 2015 The Authors
Child Development © 2015 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2015/8605-0021
DOI: 10.1111/cdev.12404

experiences (Leavens & Bard, 2011; Lyn, 2010; Pedersen, Segerdahl, & Fields, 2009). Typically, the apes tested are captive-born animals, sometimes orphaned from near birth, and their early rearing histories bear little resemblance to those of the humans with whom they are being directly compared (e.g., Bard, Bakeman, Boysen, & Leavens, 2014; Leavens, 2014; Leavens, Hopkins, & Bard, 2008). Without exception, to our knowledge, every published claim of a species difference between apes and humans in the display and comprehension of simple, deictic signals, like pointing and directed gaze, has compared culturally embedded humans with apes who have been isolated from human cultural conventions (Bard & Leavens, 2014; Boesch, 2007, 2012; Leavens, 2014; Leavens & Bard, 2011; Lyn, 2010). Moreover, to our knowledge, no published claim of a species difference in social cognition between apes and humans has ever administered the same procedure to the humans and to the apes. Therefore, if empirical studies supporting claims of human uniqueness in social-cognition comprised the only class of scientific evidence available on the question of comparative social cognition, we would be unable to judge whether these alleged differences between humans and apes in sociocognitive skill were attributable to (a) their different evolutionary histories, (b) their different life history stages, (c) their different rearing environments, (d) their different experimental protocols, (e) other systematic differences in the life experiences of captive apes and typically developing humans, or (f) some combination of these factors.

Interestingly, when apes and humans are matched on some of these chronic confounding variables, no species differences emerge (Leavens, 2004, 2014). For example, with respect to matching apes and humans on age (or life history stage), Horowitz (2003) tested human adults on an imitation task and found that they, like adult chimpanzees, tended not to precisely mimic human demonstrators, but to use idiosyncratic means for opening an "artificial fruit" (see Whiten, Custance, Gómez, Teixidor, & Bard, 1996, for a description of this apparatus and procedure; also see Froese & Leavens, 2014, for a new theory of imitation based on these findings). Thomas, Murphy, Pitt, Rivers, and Leavens (2008) exposed human adults to an experimental condition in which the participants were required to find hidden rewards in one of two possible locations, based, in one condition, on an experimenter's gaze to a fixation point in the same hemispace, but directed above, not directly at, a baited container. In the original study by Povinelli et al. (1999), adolescent chimpanzees performed above chance in this condition, whereas human 3-year-old children performed at chance (i.e., the chimpanzees were more successful than the children). They interpreted this "species difference" as evidence that the human children, due to their sophisticated reasoning processes, could not relate the experimenter's gaze to a location at which the experimenter was not directly looking. Thomas et al. found that young human adults, like the chimpanzees in Povinelli et al., easily used the experimenter's cue to find the hidden rewards, despite its focus away from the baited container. In a direct comparison of cognitive development between a human child and a chimpanzee, Kellogg and Kellogg (1933) reported broad similarity between the two subjects in, for example, the initial development of language comprehension, although the child eventually began to rapidly outpace the chimpanzee in this regard (it should be noted that the human child was 2.5 months older than the ape). Thus, when apes are approximately matched with humans on age, they tend to act more similarly, particularly in the infancy period.

With respect to matching apes with humans on pre-experimental exposure to human cultural conventions, Carpenter, Tomasello, and Savage-Rumbaugh (1995) reported that enculturated great apes (a chimpanzee, *Pan troglodytes*, and two bonobos, *Pan paniscus*) overlapped with 18-month-old human children in mean durations of joint attention episodes, although the animals were not matched to the children on age or procedure. Although the human children communicated declaratively much more than the apes, two of the three enculturated apes displayed declarative signals, whereas none of the three nonenculturated apes did (see Carpenter & Call, 2013, for an apparent retraction of this claim of protodeclarative signaling). Similarly, Lyn, Russell, and Hopkins (2010) found that enculturated apes significantly outperformed nonenculturated apes in their ability to use human pointing cues to find food. Here, the approximation to a human rearing environment produced humanlike levels of performance (see Leavens, 2014, for additional analysis of this finding). Russell, Lyn, Schaeffer, and Hopkins (2011) matched 20 great apes on their species, their ages, and, with exception of one pair of apes, on sex, and compared the performances of the 10 enculturated apes against the performances of the 10 nonenculturated animals. They found that the enculturated apes significantly outperformed the nonenculturated apes on tests of both physical and social cognition, derived from the Primate Cognitive Test Battery used by Herrmann et al. (2007).

Moreover, the enculturated apes displayed similar response profiles to the 3-year-old German children used in the original study by Herrmann et al. (2007). Again, approximate matching for pre-experimental experience between apes and humans rendered broadly similar response patterns to the same kinds of cognitive challenges across the two species.

Recently, van der Goot, Tomasello, and Liszkowski (2014) reported that great apes housed at Leipzig Zoo never communicated about food from a distance. They presented a *lean interpretation* of imperative (requestive) pointing, in which the pointing gesture is seen as a kind of abbreviated reach. Given that most children tested in laboratory contexts are restrained in some way (e.g., strapped into high chairs), the lean interpretation of imperative pointing predicts that were the physical constraints removed, the children would simply approach the objects they desire. In contrast, under a *rich interpretation* of imperative pointing, children understand that their gestures influence the attentional focus of an interlocutor. The rich interpretation of imperative pointing predicts that when constraints to locomotion are removed, children who understand the implications of their signaling for the attentional foci of their interlocuters will simply remain in place, relying on the attention-directing properties of imperative, deictic signals to achieve their goal. They applied a similar reasoning to the case of great apes who were constrained by cage mesh—if apes had an appreciation of the attention-directing characteristics of their own signaling attempts, then they should, at least sometimes, signal from a distance, rather than approaching the cage mesh. Ten of the 12 apes in their study approached a location in which both food and an experimenter were situated, and requested the food with gestures from this position, and none of the apes requested the food with gestures from further back in their enclosures, which van der Goot et al. interpreted as evidence that the apes did not discern the attention-directing aspect of their imperative signaling. In contrast, about half of the human 12-month-old infants in their study pointed without locomoting toward their rewards (toys), and in accordance with their premises, this was taken as evidence that roughly half of the human children in their study appreciated the attention-directing attribute of their imperative signaling.

van der Goot et al. (2014) argued that this apparent difference between the apes and the humans signified the possession, by the 12-month-old humans, of a sense of common ground with the experimenters. They claimed that the infants' positional behaviors implicated young infants' appreciation of these requesting contexts as fundamentally collaborative activities "where the motivation to cooperate becomes as important as the motivation to reach an individual goal" (van der Goot et al., 2014, p. 454). Thus, according to van der Goot et al., gesturing at a distance, or *distal* communication, is indicative of the appreciation of common ground in the signaler, whereas *proximal* communication is diagnostic of the absence of this appreciation of common ground.

Several aspects of this argument seem problematic to us. First, as van der Goot et al. (2014, p. 454) noted, it is not unreasonable to suppose that the appreciation of common ground must rely, developmentally, on having experiences of collaborative and responsive responses from one's caregivers. Would a child become as reliant upon the helpful motivations of their interlocutors if their experiences had not fostered those expectations? We believe that it remains ambiguous whether the subset of children who signaled without approaching their desired toys, in their study, did so because they have phylogenetic histories in which an appreciation of imperative signals as having attention-directing properties led to differential survival or reproductive success (as van der Goot et al., 2014, argued) or, alternatively, because they have had sufficient developmental experiences to foster an expectation that their signals will marshal responsive behavior from their caregivers. Conversely, it is ambiguous whether the apes in that study universally approached the referent (or experimenter) before signaling because of their evolutionary histories or because of their developmental experiences—it is implausible that these institutionalized apes have experienced the same level of responsiveness to their imperative signals from human interlocutors that the human children have experienced. This issue is of central importance to researchers interested in the origins of communicative behavior in humans and their nearest living relatives: If apes raised in institutionalized settings respond differently to experimental challenge from human children raised in Western, postindustrial settings, then because their pre-experimental histories differ so radically, it is logically unclear whether that difference is due to differences in phylogeny, ontogeny, or interactions between genes and environment (Bard & Leavens, 2014; Boesch, 2012; Leavens & Bard, 2011; Leavens et al., 2008; Racine, Leavens, Susswein, & Wereha, 2008). Thus, on deductive grounds, the attribution to phylogenetic factors by van der Goot et al. of the different performances of the apes and the humans in their study cannot withstand the criticism that the apes and the human children were not matched on pre-experimental experiences of common ground with human interlocutors. Second, the argument presented by van der Goot et al. is tautological: Had the apes in that study behaved like the human children, then there would have been no empirical basis for their claim that a psychological state that they termed "appreciation of common ground" exists, let alone that it influenced the locomotor behavior of those children. Finally, an obvious weakness of the argument by van der Goot et al. is their implication that because the 10 apes who signaled in their study all approached the experimenter when they gestured this is a general response characteristic of great apes.

Because we, and others, have repeatedly observed chimpanzees pointing to objects from positions in their enclosures that were relatively distant from those objects, we had reason to further explore whether van der Goot et al.'s (2014) findings are representative of this species. For example, distal pointing by bonobos (*Pan paniscus*) was reported by Pedersen et al. (2009) and by Veà and Sabater-Pi (1998). In addition, it has been well demonstrated that small changes to a procedure can yield substantial differences in results when studying human and nonhuman primates (e.g., Hopkins, Russell, McIntyre, & Leavens, 2013; Lyn, 2010; Lyn et al., 2014; Mulcahy & Hedge, 2012). With only 10 apes responding in their study, it remains open to question whether their findings are generalizable beyond their subset of animals from the Leipzig Zoo. Therefore, we designed a study to determine whether chimpanzees would always approach desirable food before signaling about it, as the findings of van der Goot et al. would predict, or would, instead, display the kind of mixed pattern of results displayed by the human children in their study, in which some children approached the reward, but other children did not.

We did not attempt to replicate the design of van der Goot et al. (2014) for several reasons. Their human infants were positioned at systematically closer distances to the reward location (0.95 and 1.8 m, in two different experiments), compared to their great apes, who were separated from the reward location by 6 m. Any attempt to replicate the study, therefore, would have to systematically vary the task administration across these three distances, because these distance and species effects were confounded factors in the original study (i.e., there is no objective basis to decide whether the reported behavioral differences between the apes and the humans were attributable to the organisms' different evolutionary histories or the radically different physical arrangements). Similarly, the human children in van der Goot et al. were sitting next to the experimenter with no barrier between them, whereas the great apes were separated from the human experimenters by cage mesh. The orientations and positions of the subjects and the two experimenters varied substantially and systematically with species classification, so, again, to replicate the study would have required flexibility in the physical testing environments that was not available to us with our study population. Finally, in the experimental arrangement administered by van der Goot et al., an experimenter was standing adjacent to the target location; hence, it is ambiguous to which of the two experimenters the apes were signaling.

Therefore, we adopted a new procedure, based on established protocols in our laboratory (Leavens & Hopkins, 1998, 2005; Leavens, Hopkins, & Thomas, 2004), designed to render a clear answer to the question, "Do chimpanzees communicate about food from a distance, or do they always approach as closely as they can to the food, before signaling?" Our general tactic was to (a) reduce the number of potential communicative partners (experimenters) to one, to reduce ambiguity about the intended recipient; (b) separate the referent from the communicative partner with a wide angle, to reduce ambiguity about whether deictic signals were directed at a person or an object; and (c) approximately match the distance between subjects and the referents with those used with the human children in van der Goot et al. (2014), to eliminate distance as a confounding factor. We presented four trials to each of 166 chimpanzees in an experimental situation in which an experimenter and desirable food were presented to the animals at significant angular displacement, so that it was never ambiguous whether their manual gestures and other signals were food directed or person directed.

We reasoned that if van der Goot et al. (2014) are correct in their claim that the appreciation of imperative signaling as an attention-directing device is a uniquely human capacity, then all of the chimpanzees should approach the food before signaling about it to an interlocutor, consistent with their lean interpretation of imperative signaling as a psychologically simple, instrumental act.

For researchers interested in child development, if we find that chimpanzees, like human children, display a mixed pattern of response—some communicating at a distance, and some moving as close as possible to the food—then this would imply that there are not substantive psychological differences between chimpanzees and younger human children in their imperative communication. This conclusion would hold regardless of the plausibility of the argument that distal signaling implicates the understanding of imperative communication as an attention-directing signal.

Method

Subjects

In total, 166 chimpanzees were tested in two colonies. Subjects were 125 subadult and adult chimpanzees housed at the University of Texas MD Anderson Cancer Center (UTMDACC) ranging in age from 13 to 52 years of age (M = 31.6 years, SD = 9.8), tested between January and June 2014. There were 74 females and 51 males. Within the sample, there were 71 mother-reared, 20 human-reared, and 34 wild-caught apes. Subjects in the sample from Yerkes National Primate Research Center (YNPRC) were 41 subadult and adult chimpanzees ranging from 10 to 58 years of age (M = 28.2 years, SD = 11.0), originally filmed between May 2011 and March 2013; these videos were recoded for the present study between July and November 2014. There were 29 females and 12 males. In this sample, there were 19 mother-reared, 19 human-reared, and three wild-caught apes. On average, our sample was older (combined M = 30.7 years, SD = 10.2) than the mean 17.7 years reported for the apes in van der Goot et al. (2014), but our range of 10–58 years considerably overlaps their range of 9–35 years. Systematic analyses revealed no differences in propensity to display either initiation of joint attention (IJA; defined below) or in triangulated signaling (TRI; defined below) as a function of age class (see Results); therefore, any differences that emerge between our sample and that of van der Goot et al. are unlikely to be attributable to age. Like apes in most captive circumstances, the UTMDACC and YNPRC chimpanzees were fed 2–3 times daily with commercial primate chow, liberally supplemented with fresh fruits and vegetables. Unlike zoo apes, chimpanzees housed in research facilities are not put out on display during the day, but like zoo apes many have night enclosures in which they sleep, ranging in larger corrals during the daytime. Of particular note, like the apes in van der Goot et al., the UTMDACC and YNPRC chimpanzees were significantly older than the human infants in their study, yet, like the human infants in van der Goot et al., and human infants more generally, they frequently encounter desirable items that are outside their reach and develop tactics for controlling the activities of caregivers apparently to obtain those otherwise unreachable items, for example, through pointing (e.g., Leavens, Hopkins, & Bard, 1996, 2005; Leavens et al., 2008).

Procedure

Prior to testing, each subject was encouraged to separate from their usual social groups, either into their home cages or in familiar runs, depending on the specific circumstances. Testing areas ranged from 5.76 to 13.38 m^2. Interactions between these chimpanzees and humans, including routine feeding and husbandry, as well as experimental manipulations, typically take place through the cage mesh in these areas of the animals' captive environments; hence, the distances involved in the present study (see below) are more typical of the interactive distances involved in the day-to-day interactions of chimpanzees than the 6.0 m used in van der Goot et al. (2014). Immediately prior to each trial a human experimenter (E1) was positioned outside the cage mesh at one end of the subject's home cage. After a short period of time, a second experimenter (E2) approached the enclosure and placed a container of preferred food 0.5 m outside the subject's home cage, at the end opposite of the position of E1, in full view of the chimpanzee, then E2 immediately left the area. The distance between E1 and the food was approximately 1–1.5 m, and the placement of the food signaled the onset of the trial.

During the next 60 s, E1 recorded whether or not the chimpanzee signaled to either them or the food and alternated their gaze between the food and E1. Gaze alternation between a social partner and a referent is a widely used and easily operationalized component of intentional communication in humans (Bates, Benigni, Bretherton, Camaioni, & Volterra, 1977; Desrochers, Morissette, & Ricard, 1995; Franco & Butterworth, 1996; Lempers, 1979; Tomasello, 1995), and is used to index the social orientation of accompanying communicative behavior (see, e.g., Bard, 1992; Leavens & Hopkins, 1999; Leavens, Russell, & Hopkins, 2005; Plooij, 1978, for discussion of its extension to great ape communication). (Although some contemporary researchers write as though such behavioral indices as gaze alternating behavior provide insight into psychological mechanisms, this interpretation is rejected by Leavens [2012; e.g., Leavens, Racine, & Hopkins, 2009; Leavens, Russell, et al., 2005; Racine et al., 2008]; here, we simply note, following Leavens, Russell, et al. [2005] that gaze-alternating behavior meets an objective behavioral criterion for conventional definitions of intentional communication in both apes and humans.) We used an expanded ethogram, compared to van der Goot et al., because we reasoned that when a chimpanzee stands in front of a human experimenter, directs its hand in a begging posture toward the experimenter, and alternates its gaze between this experimenter and a distant, desirable referent, this is classically triadic communication about a particular entity, and therefore the propensity of these animals to communicate about this food would be underestimated if we arbitrarily recorded only a subset of their triadic communicative repertoires (these deictic repertoires have been described in Leavens & Hopkins, 1998; Leavens et al., 2004; Leavens, Russell, et al., 2005). When the chimpanzees signaled with gaze alternation, this was recorded as a successful IJA response (e.g., Mundy, Kasari, & Sigman, 1992). Signaling responses included pointing, cage banging, pushing an object toward the human or food in an apparent barter attempt (Hyatt & Hopkins, 1998), spitting (rare, only two occasions), or throwing (also rare, only one occasion). In the absence of gaze alternation between the human and the food, presenting a body part or using the feet to signal were not considered communicative responses for the purposes of this study. Note that the chimpanzee had to be simultaneously signaling to either the food or human while alternating gaze between the two items. These events could not occur in sequence to be considered a successful joint attention event. In other words, signaling to the food or E1, retracting their gesture or stopping their response, and then alternating their gaze was not considered a successful IJA response. In addition, E1 also noted the position of the chimpanzee when they initiated joint attention as either (a) in front of the food (food) or (b) in front of the human (experimenter). In relation to the coding scheme of van der Goot et al., like them, we included points, but unlike them, because we separated the food and the experimenter, we had two classes of manual extensions or gestures: (a) point (to food) and (b) hold hand out (apparently supplicative gestures directed at the experimenter; see Leavens & Hopkins, 1998; Leavens et al., 2004). van der Goot et al. reported grabbing of the cage mesh, but we did not see this behavior in the present study. Behaviors coded by us, but not by van der Goot et al., included cage bangs, bartering attempts, spitting (two instances), and throwing (one instance).

In addition to IJA, we computed whether the chimpanzees engaged in the more restricted form of IJA that we defined as TRI. TRI was defined as the chimpanzee positioning themselves either in front of the human or food and signaling to the more distal referent, be it the human or food. Thus, in these particular cases of IJA, the chimpanzees would either position themselves in front of the food and signal to E1 (and not the food) or they would position themselves adjacent to E1 and signal toward the food (and not E1). Lastly, on some trials, the chimpanzees positioned themselves between the food and E1 and signaled to both items while alternating their gaze. These responses were computed as a separate form of triangulation. Thus, TRI included all instances of IJA in which the chimpanzees' manual gestures were directed to the relevant target (food or E1) that was furthest away from its position, plus the few trials in which the chimpanzees positioned themselves between the food and E1 and signaled to both targets.

Each subject received four trials that lasted 60 s each. There was at least a 4-min interval between successive trials. All trials were videorecorded at both colonies. There were some slight differences in coding procedure between UTMDACC and YNPRC: At UTMDACC, data were recorded live on datasheets by E1 and the videotaped record consulted to resolve uncertainties, when they arose. At YNPRC, all trials were videorecorded and then subsequently coded from these video records by the fourth and fifth authors of this article. Though the subjects lived in social groups, they were temporarily separated from their groups during testing, as were the apes in van der Goot et al. (2014), and tested in their home cages. The apes' separation from their social groups was voluntary, and trials were to be terminated in the event of any apparent distress, in accordance with our usual policies, but no subjects displayed any apparent distress in this procedure. No attempt was made to record or manipulate the chimpanzees' initial positions in their cages. E1 and E2 were randomized across subjects and trials; these roles were performed by the second and third authors of this article at UTMDACC and by the fourth and fifth authors at YNPRC.

Analyses and Reliability

All statistical tests were two-tailed, with alpha set to .05; chi-square tests, Fisher's exact test, Kruskal–Wallis tests, and Mann–Whitney tests were used throughout, consistent with the nominal level of measurement. Prior to testing, interrater reliability was established using Cohen's kappa. Ten chimpanzees were filmed while being tested and their four-trial performance was independently scored for IJA by these two independent raters, the second and third authors of this article, who were blind to the hypotheses under test during both filming and later data collection (*n* = 40 observations, 4 trials × 10 chimpanzees) and who served as E1 and E2. These 10 chimpanzees were later tested again in the formal testing protocol. The percentage agreement in IJA ratings was 92.1% ($\kappa = .842$), which is considered excellent (Bakeman & Quera, 2011). Note that TRI classifications were computed from the database recording the behavioral constituents of IJA responses (i.e., all instances in which the gestures were directed to the relevant target furthest away from the closest target at which the chimpanzees were positioned were coded as TRI).

Results

Descriptive Performance

In the initial set of analyses, we examined whether there were any systematic colony-based differences in responses. For these analyses, we used the number of trials on which the chimpanzees displayed IJA or a triangulated (TRI) response. There were no differences between the UTMDACC (*n* = 125) and YNPRC (*n* = 41) colonies on the propensity to display IJA (*U* = 2469.5, *p* = .601; UTMDACC: *Mdn* = 2.0, YNPRC: *Mdn* = 2.0) or TRI (*U* = 2388.5, *p* = .491; UTMDACC: *Mdn* = 1.0, YNPRC: *Mdn* = 1.0). In addition, we checked for differences between the two colonies in positional behavior (proximal vs. distal positioning) on a trial-by-trial basis. We found no systematic differences in positional behavior, where subjects who displayed IJA were categorized as having taken up positions proximal to the food, or not, between UTMDACC and YNPRC: Trial 1, $\chi^2(1, N = 82) = 0.02, p = .894$; Trial 2, $\chi^2(1, N = 98) = 2.11, p = .146$; Trial 3, $\chi^2(1, N = 91) = 3.51, p = .061$; and Trial 4, $\chi^2(1, N = 92) = 0.96, p = .328$. Therefore, we combined data across the two colonies for all subsequent analyses.

In a second set of initial analyses, we examined the potential influences of sex and rearing history on performance. The trial-specific and overall data are shown in Tables 1 and 2. Mann–Whitney *U* tests were used to test the effects of sex, and Kruskal–Wallis tests were used to test for effects of rearing history on the number of trials with IJA and TRI respectively. No significant effects of sex (*n* = 103 females, *n* = 63 males) on overall IJA responding were found (Mann–Whitney *U* = 3167.5, *p* = .793; females: *Mdn* = 2.0, males: *Mdn* = 2.0). No significant effects of rearing history were found on IJA performance, Kruskal–Wallis *H*(2, *N* = 166) = 4.63, *p* = .099 (mother reared:

Table 1

Number of Chimpanzees With IJA Responses by Rearing History, Sex, and Number of Trials

	Number of correct IJA responses					
	0	1	2	3	4	Percent correct
Rearing						
Mother (*n* = 90)						
Female (*n* = 56)	14	12	9	8	13	47.3
Male (*n* = 34)	7	7	5	6	9	52.2
Nursery (*n* = 39)						
Female (*n* = 23)	4	1	3	7	8	62.3
Male (*n* = 16)	2	3	3	3	5	59.4
Wild (*n* = 37)						
Female (*n* = 24)	4	3	6	2	9	59.4
Male (*n* = 13)	1	3	3	0	6	63.5
Overall (*N* = 166)	32	29	29	26	50	57.8

Note. For each row, the number of subjects is the total number of trials from which the percentages are derived. IJA = initiation of joint attention; Percent correct = (the number of correct trials divided by the total number of trials [4]) × 100.

Table 2

Number of Chimpanzees With TRI Responses by Rearing History, Sex, and Number of Trials

	Number of correct TRI responses					
	0	1	2	3	4	Percent correct
Rearing						
Mother (*n* = 90)						
Female (*n* = 56)	28	13	9	3	3	23.2
Male (*n* = 34)	16	8	1	4	5	30.9
Nursery (*n* = 39)						
Female (*n* = 23)	7	4	6	6	0	37.0
Male (*n* = 16)	6	4	2	2	2	34.4
Wild (*n* = 37)						
Female (*n* = 24)	12	4	4	1	3	28.1
Male (*n* = 13)	5	4	0	1	3	36.5
Overall (*N* = 166)	74	37	22	17	16	31.7

Note. For each row, the number of subjects is the total number of trials from which the percentages are derived. TRI = triangulated signaling; Percent correct = (the number of correct trials divided by the total number of trials [4]) × 100.

Mdn = 2.0, nursery reared: *Mdn* = 3.0, wild born: *Mdn* = 2.0). Similarly, no significant effects of sex on TRI performance were found (Mann–Whitney *U* = 3058.0, *p* = .512; females: *Mdn* = 1.0, males: *Mdn* = 1.0). Finally, no significant effects of rearing history were found on TRI performance, Kruskal–Wallis *H*(2, *N* = 166) = 3.04, *p* = .219 (mother reared: *Mdn* = 1.0, nursery reared: *Mdn* = 1.0, wild born: *Mdn* = 1.0). Given the apparent absence of sex or rearing history effects on either IJA or TRI performance, the data were pooled across these subject variables for comparison with van der Goot et al. (2014).

In a third set of preliminary analyses, we examined the influence of subjects' ages on IJA and TRI performance. Although the UTMDACC colony was, on average, about 3 years older than the YNPRC sample (see Method), this difference was not statistically significant, *t*(164) = 1.86, *p* = .065; therefore, the two colonies were combined for further analysis. We categorized each chimpanzee as belonging to one of four age classes: (a) adolescent (8–12 years, *n* = 2), (b) young adult (13–20 years, *n* = 19), (c) mature adult (21–35 years, *n* = 99), and (d) elderly (> 35 years, *n* = 46); these age classes correspond to those used by Leavens and Hopkins (1998). Systematic chi-square analyses performed for each of the four trials compared these four levels of age class with two levels each of, respectively, IJA and TRI performance (i.e., the chimpanzees either displayed IJA or not, on any given trial, and either displayed TRI or not, on any given trial). There were no systematic effects of age on IJA or TRI performance—IJA: Trial 1, $\chi^2(3, N = 166) = 5.12$, $p = .163$; Trial 2, $\chi^2(3, N = 166) = 0.46$, $p = .926$; Trial 3, $\chi^2(3, N = 166) = 1.64$, $p = .650$; Trial 4, $\chi^2(3, N = 166) = 3.64$, $p = .302$; and TRI: Trial 1, $\chi^2(3, N = 166) = 1.36$, $p = .716$; Trial 2, $\chi^2(3, N = 166) = 0.42$, $p = .935$; Trial 3, $\chi^2(3, N = 166) = 1.72$, $p = .633$; Trial 4, $\chi^2(3, N = 166) = 0.98$, $p = .807$.

As is evident in Table 1, on average, chimpanzees initiated joint attention on 57.8% of the trials with 30.1% producing IJA on all four trials (50/166) compared to only 19.3% who never displayed IJA (32/166). A triangulated response was observed less frequently, with chimpanzees, on average, producing this form of joint attention signaling on 31.7% of trials. Of the 166 chimpanzees, only 16 individuals (9.6%) used this form of communication on all four trials compared to 74 individuals (44.6%) who never used it (see Table 2).

A reviewer asked whether frequent communicators (IJA on three to four trials, *n* = 76) positioned themselves differently with respect to less frequent communicators (IJA on one to two trials, *n* = 58). In other words, could it be the case that positioning preferences and propensity to communicate are confounded, in the present study? To answer this question, we calculated an ordinal variable to describe the distance at which the chimpanzees displayed an IJA response. This was the sum of the total distances recorded over all four trials, divided by the number of trials in which subjects displayed IJA. Subjects who positioned themselves at the food on any given trial were coded as 1, subjects who positioned themselves midway between the food and the experimenter on any given trial were coded as 2, and subjects who positioned themselves near the human and communicated most distally were coded as 3 (representing the most distal position). The sums over the four trials could range, thus, from 1 (*a subject who positioned itself at the food on only one trial and did not otherwise display IJA*) to 12 (*a subject who positioned itself near the human and displayed IJA on all four trials*). This sum was divided by the number of trials in which each subject displayed IJA, to derive a mean ordinal distance. A Mann–Whitney *U* test comparing infrequent responders (IJA on one to two trials) with frequent responders (IJA on three to four trials) was not significant (*U* = 1913.5, *p* = .178; infrequent IJA: *Mdn* = 1.00, frequent IJA: *Mdn* = 1.7). Thus, it was not the case that frequent IJA responders displayed a systematically different positioning tactic, compared to less frequent IJA responders.

Gesture Types

Table 3 reports the frequencies and percentages of communicative acts displayed by these chimpanzees, by trial number. For relatively rare multiple signal types, here we classified communication acts in accordance with the following hierarchy: gesture to food > gesture to experimenter > other (e.g., if a

Table 3

Frequencies (and Percents) of Communicative Acts by Type

	Trial number			
	1	2	3	4
No communication	84 (51)	68 (41)	75 (45)	74 (45)
Gesture to food (pointing)	32 (19)	38 (23)	43 (26)	48 (29)
Gesture to experimenter	37 (22)	45 (27)	33 (20)	30 (18)
Other	13 (8)	15 (9)	15 (9)	14 (8)

Note. *N* = 166. "No communication" includes apparent lack of signaling, exhibitions of parts of the body, plus apparent signals not accompanied by gaze alternation. "Other" includes cage bangs, bartering attempts, spitting, and throwing. Percentages may not sum to 100 due to rounding error.

subject pointed at the experimenter, then pointed at the food, this would be classified as a gesture to the food). On any given trial, between 41% and 50% of the chimpanzees displayed a manual gesture directed toward the food or the experimenter. Of these, between 46% and 62% were pointing gestures. Other, nondeictic communicative acts comprised between 8% and 9% of the responses across the four trials. Between 19% and 29% of the chimpanzees pointed to the food, whereas between 41% and 51% of the chimpanzees displayed "no communication" (which includes apparently communicative signals that were not accompanied by gaze alternation between the food and the experimenter, as well as trials in which the chimpanzees did not engage with the experimenter). Photographic examples of some of the gestures are included as Figures S1–S5 in the online Supporting Information.

Signaling Positional Effects

To test the hypothesis that chimpanzees always communicated from the position closest to the food, we compared the occurrence of IJA responses as a function of whether the chimpanzee was positioned in front of the food (food), in front of the human (experimenter), or between these two entities (middle) using chi-square tests for each trial. Some individuals changed position in the course of a trial; we coded all subjects in terms of their first signaling position for these analyses. These data are shown in Figure 1. As is evident from Figure 1, there was a substantial preference displayed by these chimpanzees for communicating either in closest proximity to the food or in proximity to the experimenter, with relatively fewer animals electing to signal from a position midway between the food and the experimenter: Trial 1, $\chi^2(2, N = 82) = 27.24$, $p < .001$; Trial 2, $\chi^2(2, N = 98) = 23.41$, $p < .001$; Trial 3, $\chi^2(2, N = 92) = 8.86$, $p = .012$; and Trial 4, $\chi^2(2, N = 93) = 19.39$, $p < .001$. However, when the two distal positions are combined (experimenter + middle = distal) and compared to the individuals who moved to proximity with the food (= proximal), systematic binomial tests reveal no bias toward either proximal or distal signaling on any trial (binomial tests, Trial 1: $p = .097$, Trial 2: $p = 1.000$, Trial 3: $p = .142$, Trial 4: $p = 1.000$; see Figure 2). (The rationale for combining middle and experimenter signalers for this analysis is that both responses meet the criterion of not approaching as closely as possible to the food.) Because

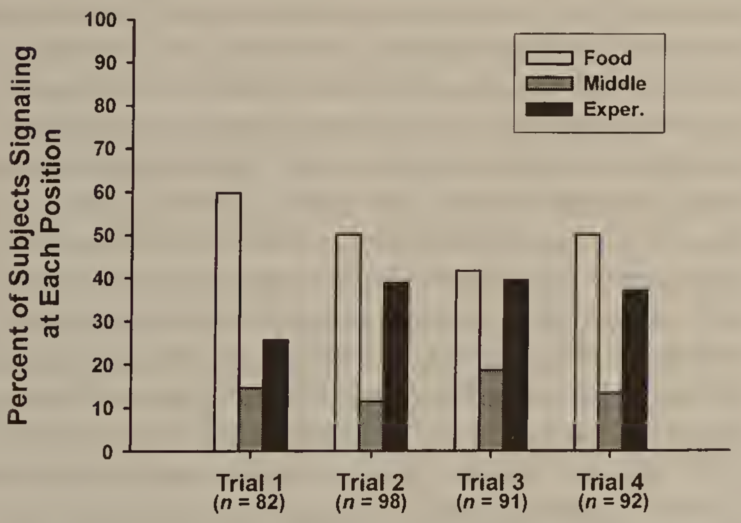


Figure 1. The positional behavior of the chimpanzees as the percent of subjects who (a) approached the food and communicated with the experimenter (Food), (b) positioned themselves between the food and the experimenter and communicated about the food with the experimenter (Middle), and (c) approached the experimenter and communicated about the food with the experimenter (Exper.). Sample sizes refer to the number of individuals who signaled out of the total sample of 166 chimpanzees.

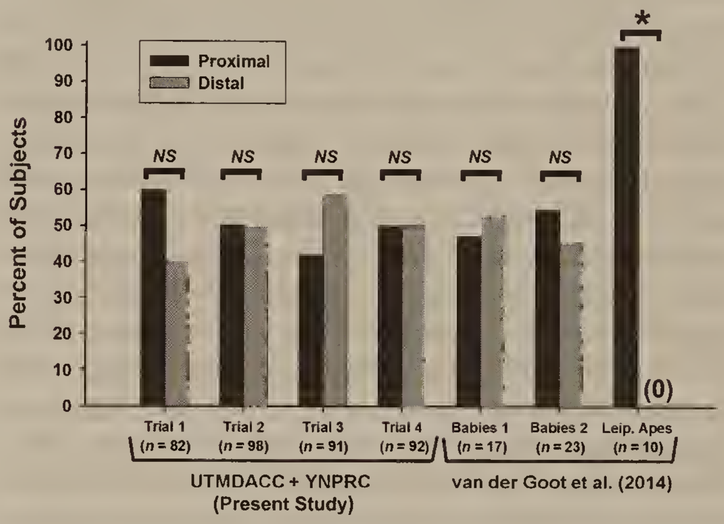


Figure 2. The percent of subjects who communicated from a position proximal to the food (Proximal) and the percent of apes who communicated from a position distal to the food (Distal) in the present study (Trials 1–4, chimpanzees at University of Texas MD Anderson Cancer Center [UTMDACC] and Yerkes National Primate Research Center [YNPRC]). Babies 1 and Babies 2 refer to the human infants in van der Goot et al. (2014), their Experiments 1 and 3, respectively. The last group on the right are the apes in the original study at the Leipzig Zoo by van der Goot et al. (Leip. Apes). Chimpanzees who communicated from the Middle or near the Experimenter were combined and categorized as Distal communicators for the purposes of this comparison. Sample sizes refer to the number of individuals who signaled of the total sample of 166 chimpanzees (left to right, the first four trials from the present study comprising chimpanzees from UTMDACC and YNPRC), the total number of babies (Babies 1 and Babies 2) who signaled or locomoted in Experiments 1 and 3, respectively, from van der Goot et al., and, in the last group, the apes at Leipzig Zoo, from van der Goot et al. Statements about significance are from the respective studies (UTMDACC and YNPRC chimpanzees, Trials 1–4 from the present study; Babies 1, Babies 2, and Leip. Apes from van der Goot et al., 2014); data for the last three groups from van der Goot et al. (their Figure 2). *NS* = not significantly different.
*$p < .05$.

the food is out of reach even when the chimpanzees move to the closest possible physical proximity with it, even the "proximal" signaling we report, here—using the term as van der Goot et al. did—involves approximately 50 cm of distance between the chimpanzees and the referents of their communication. Thus, in our study, the apes did not use a tactic of approaching the food either exclusively, as in van der Goot et al., or even preferentially. The percentages of individuals who were positioned distally, either in front of the human or between the human and the food for Trials 1 through 4, were 40%, 50%, 58%, and 50%, respectively. Among these animals who gestured from a distal position, the number (and percentage) of chimpanzees who pointed to the food on Trials 1 through 4 were 8 of 33 (24%), 22 of 49 (45%), 29 of 53 (55%), and 27 of 46 (59%), respectively.

Comparisons With van der Goot et al. (2014)

van der Goot et al. (2014) reported that of the 10 apes in their study who responded to the experimental challenge, none of them ($n = 0$) gestured from the backs of their cages ("All 10 apes moved as close as possible to the food items behind the cage and requested from there by protruding the hand through the mesh"; van der Goot et al., 2014, p. 450). We categorized all chimpanzees who responded during the present experiment as positioning themselves (a) in front of the food and communicating from there (proximal) or (b) either between the food and the experimenter or in front of the experimenter and communicating from there (distal). Systematic comparisons with van der Goot et al.'s data reveal that the present sample was significantly more likely than the Leipzig apes to communicate from a position that was not close to the food for all trials (see Figure 2): Trial 1 (Fisher's exact test, $p = .012$), Trial 2 (Fisher's exact test, $p = .002$), Trial 3 (Fisher's exact test, $p < .001$), and on Trial 4 (Fisher's exact test, $p = .002$). For these comparisons, we compared the 10 apes from Leipzig with 82, 98, 91, and 92 chimpanzees from the present study on Trials 1 through 4, respectively. Thus, the behavior patterns displayed by the Leipzig apes do not seem to match the UTMDACC and YNPRC apes and, therefore, are not generalizable to all populations of captive apes. Moreover, it is clear from Figure 2 that the chimpanzees we studied do not differ in any systematic way from the human babies. Comparing the positional behaviors of van der Goot et al.'s babies in their Experiment 1 ($N = 17$) with our chimpanzees in Trials 1–4 reveals no differences in Trial 1 (Fisher's exact test, $p = .421$), Trial 2 (Fisher's exact test, $p = 1.000$), Trial 3 (Fisher's exact test, $p = .791$), and Trial 4 (Fisher's exact test, $p = 1.000$). Similarly, comparing the positional behaviors of van der Goot et al.'s babies in their Experiment 3 ($N = 23$) with our chimpanzees also reveals no differences in Trial 1 (Fisher's exact test, $p = .808$), Trial 2 (Fisher's exact test, $p = .815$), Trial 3 (Fisher's exact test, $p = .341$), and on Trial 4 (Fisher's exact test, $p = .814$). As is evident from Figure 2, the Leipzig apes were the only group to display a significant preference for proximal versus distal signaling.

Discussion

In contrast to the chimpanzees and bonobos studied by van der Goot et al. (2014), the chimpanzees at UTMDACC and YNPRC frequently communicated about the food from positions that were not the closest possible position to the food; that is, they displayed distal deictic communication. This is consistent with casual observations conducted during our previous research. It is precisely this pattern of not approaching the target as closely as possible by the human children that led van der Goot et al. to argue that "infants' imperative pointing is a cooperative communicative act that directs others' attention within common ground to desired objects, with the expectation that the recipient will do as one wishes" (p. 448). Our finding that chimpanzees frequently do this has considerable relevance to the theoretical claims of van der Goot et al. Recall that they argued that this kind of distal signaling was (a) unique to humans and (b) diagnostic of the capacity to recognize a common, shared psychological stance in others. With respect to the first claim, our findings clearly demonstrate that distal signaling is not a uniquely human communicative tactic. Across the four trials, anywhere from 38% to 63% of the responding apes chose to communicate from a position other than at the closest approach to the desirable food. We do not know why our sample of great apes differs so much from the Leipzig group, but we note that the distances involved in our study (~1.0 to ~1.5 m) were much closer to those used with the human children (0.95 to 1.8 m) than the distances administered to the great apes in van der Goot et al.'s study (~6 m), and our chimpanzees responded very similarly to both groups of human children in van der Goot et al. in the display of distal signaling (see Figure 2). This raises the possibility that simple distance may be among the most influential factors controlling the positional concomitants to signaling behavior in both humans and apes. In addition, and in contrast to van der Goot et al., we did not confound the location of an experimenter and the desired reward. Hence, perhaps it was the presence of the experimenter (E2, in their study) that drew the Leipzig apes so consistently toward the food in their study.

With respect to the second claim, that distal signaling constitutes an index of the capacity of organisms to recognize a state of shared psychological perspective, or common ground, must we now concede this ability to great apes? According to van der Goot et al. (2014), this is precisely the interpretation we should deliver about the chimpanzees at UTMDACC and YNPRC:

> The fact that they always move their body as close as possible toward the desired item suggests that apes are using a strategy of enacting their message rather than presuming the message to be within common ground so that directing attention from afar would suffice. (p. 450)

Thus, according to van der Goot et al. (2014), distal signaling indexes the presumption that one's signals are understood within a "common ground" that is shared between both signaler and receiver. It is worth noting that the fact that the children in van der Goot et al. often approached the reward, indeed, in proportions comparable to the chimpanzees in the present study (Figure 2), must be construed, based on their argument, as evidence that these children lacked a sense of the common ground in which the experiment was conducted. According to the claim of van der Goot et al., organisms who approach the reward in these triangular experimental contexts (like the majority of their children and like many of our chimpanzees) lack a sense of common ground. This proximal signaling is their index of the absence of common ground in communicative interactions. Conversely, their behavioral index of the psychological appreciation of a common ground is the distal signaling displayed by many of their children and many of our chimpanzees.

An alternative, arguably simpler explanation for the mixed tactics of approach and nonapproach displayed by the infants in van der Goot et al. (2014) is simply that their sample of children was composed of children displaying a spectrum of social inhibition (Fox, Henderson, Marshall, Nichols, & Ghera, 2005; Garcia Coll, Kagan, & Reznick, 1984). There are well described individual differences in approach behavior among humans, so to find that some infants approached the toys, whereas others did not, as van der Goot et al. (2014) reported, is not unexpected. The striking difference between the chimpanzees in the present study and the great apes in van der Goot et al. (Figure 2) is somewhat less amenable to explanation by appeal to temperamental factors. Perhaps there may be systematic differences in the levels of perceived environmental threat between research institutions and zoos, which differentially influence the readiness of these apes to approach humans. Consistent with this conjecture, whereas 17% (2 of 12) of the apes in van der Goot et al. declined to respond to the experimental challenge, in the present study between 41% and 51% (within the four trials) of the chimpanzees declined to engage with the experiment with the display of communicative behavior (Table 3). Thus, at the very least, it seems reasonable to suggest that noncognitive temperamental and environmental factors might have influenced the patterns of response reported in these two studies.

Without necessarily speaking to the utility of the concept of a common ground in understanding organismic interactions, we argue that this concept is not well operationalized, if the argument is that organisms who display proximal communication lack a concept of common ground or, conversely, that organisms who display distal communication possess this concept.

In summary, according to the argument presented by van der Goot et al. (2014), many of our chimpanzees displayed an appreciation of the common ground between themselves and their human interlocutors; however, we are reluctant to credit these animals (or even human infants) with such a sophisticated psychological appreciation on the basis simply of where they positioned themselves while communicating about desirable entities. We believe that the present findings significantly undermine the validity of the measure of proximity in signaling as an index of the appreciation of common ground, in any species. However, there is ample room, at the present state of knowledge, and notwithstanding our skepticism, to argue that both the chimpanzees who signaled about food from a distance in the present study and the children who did so in van der Goot et al. were displaying their appreciation of a common ground. In short, either (a) van der Goot et al. have correctly identified a behavioral index of the appreciation of a common ground and our findings demonstrate that this capacity is also evinced by chimpanzees, or (b) the proximity of signaling is not a valid index of the appreciation of the state of common ground in any species.

For van der Goot et al. (2014; see also related arguments by Povinelli & Eddy, 1996; Tomasello, Carpenter, & Liszkowski, 2007) it is the fact of the differences in response patterns between humans and apes that they claim supports their interpretation of psychological differences between humans and apes in these simple experimental protocols. What our data demonstrate, with larger samples, is that there is no empirical, behavioral basis to support the claim that great apes and young humans differ in psychological responses to these kinds of simple challenge. Importantly, in our view, neither van der Goot et al. nor the present study can directly address questions about the mechanisms supporting hidden, hypothetical psychological processes, because all we measured was overt communicative and locomotor behavior, and we did not manipulate response outcomes for the subjects. Virtually all animals communicate and locomote, and there is no apparent reason to believe that merely documenting the confluence of these activities in humans or nonhumans will give us any special insight as to the cognitive processes involved—to accomplish this, at the very least, the subjects would need to be subjected to an assessment of their discriminative capacities in the face of differential response outcomes, a technique not employed either here or by van der Goot et al.

There is an additional concern that we would like to raise about the nature of species comparisons. There are several contemporary practices that, we believe, are seriously hampering theoretical development in the area of comparative-developmental evolutionary psychology. The first problem is that of overgeneralization from small samples. As demonstrated by the present study, van der Goot et al. (2014) overgeneralized from their particular sample of zoo-housed apes to all apes, everywhere. Their theoretical interpretation of the absence of distal communication in their sample requires that their subjects are representative of their species (chimpanzees and bonobos). As Figure 2 makes abundantly clear, their ape subjects' response patterns are not at all typical of captive apes; in fact, their apes behaved in an extremely unusual manner. To be clear, we have no objection to small-sample studies—what we object to is the repeated failure by practitioners to appreciate that findings in short, small-subject studies ought not to be taken as representative of the species, until confirmed with additional study. Of course, van der Goot et al. are not alone in overgeneralizing from their samples. For example, Povinelli and his colleagues (e.g., Povinelli & Eddy, 1996; Reaux, Theall, & Povinelli, 1999) claimed that chimpanzees were oblivious to the direction of human eyes, but Hostetter, Russell, Freeman, and Hopkins (2007) found that similarly institution-housed chimpanzees had no systematic difficulty discerning human eye orientations and visibility, when they were simply allowed to choose their own responses to differentially manipulated experimenters' gaze directions, and many similar refutations exist in the literature.

False negative findings in great apes have been widely promulgated when unrepresentative findings were overgeneralized and when human and ape samples were not age matched, or matched on critical rearing experience variables. These false negatives can have a disproportionate effect on theory construction in the study of human and nonhuman sociocognitive development. Particularly in an age in which we have unprecedented capacities to sample brain morphology and function in vivo and rapidly developing capacities for genomic sequencing, our discipline is vulnerable to amplified effects of false premises. For example, there is a widespread assumption that humans have unique capacities for understanding gaze or pointing or other deictic cues, which has been refuted by simple reviews of the existing literature (Leavens, 2014; Lyn, 2010; Mulcahy & Hedge, 2012). Some have argued that humans have uniquely collaborative motivations or unique capacities to form common ground with others even before they can speak, but this is significantly challenged by the findings of Volter, Rossano, and Call (2015). Again, it has been argued that humans have unique capacities to communicate about absent entities, which was refuted by Lyn et al. (2014) and, indeed, falsified in advance by virtually all of the ape language training studies since Kellogg and Kellogg (1933). Finally, it has been argued that humans have unique capacities for imitation (refuted by Custance, Whiten, & Bard, 1995; Whiten et al., 1996). The danger is that neuroscientists and geneticists will squander significant time and other resources attempting to find the neural or genetic correlates of these false species differences; because numerous, systematic differences do exist between humans and our nearest living relatives in genomes and brain anatomy and function, these will be false associations attributable to false premises. We believe that the solution lies in providing sufficient training to display the capacity in question; if an organism lacks an ability, then no amount of training will inculcate that ability. If a learning pathway exists to a competency, then it is invalid to argue that the species in question lacks that competency. In sum, having well-defined, documented phenotypes that are not subject to change with simple variations in methodology will provide a much more powerful framework for scientists interested in the genetic and neural basis of human cognitive specializations.

Finally, we think it is important to recognize that institutionalized apes, like those studied by van der Goot et al. (2014) and ourselves, necessarily live in the least complex social and physical environments of all apes, particularly in comparison to apes born and raised in the wild or to apes who have been raised as part of a human family-like environment (e.g., Bard & Leavens, 2009, 2014; Boesch, 2012; Kellogg & Kellogg, 1933; Leavens & Bard, 2011; Lyn, 2010). Nevertheless, institutional environments do pose unique ecological challenges. In particular, the physical circumstance of barriers to direct grasping of objects poses what we term the "referential problem space" (Leavens, Hopkins, et al., 2005), a repeated circumstance in which one organism is dependent upon another organism to act upon the world for them. These kinds of circumstances are exceedingly rare in wild habitats, but when these rare circumstances do occur, apparently deictic behavior is displayed by wild chimpanzees, at a distance, and with no intervening physical barriers (e.g., when one animal is dominating a food source; see Hobaiter, Leavens, & Byrne, 2014). Thus, we interpret apes' communication about distal entities as part of a general tactic of social manipulation that is also displayed by human children who spend enormous amounts of time in conditions of physical restraint (e.g., in car seats, cribs, cots, backpacks, feeding chairs, etc.), leaving them reliant upon others to act on the world for them (Leavens, Hopkins, et al., 2005). Because tactics of manipulation with deictic signals emerge in these kinds of circumstances in both humans and apes, we believe that there is no evidence-based rationale to conclude either that (a) humans have unique cognitive capacities to support their deictic signaling or (b) deictic signaling is, necessarily, a cognitively advanced ability. There may be, however, uniquely human cognitive processes associated with distal pointing, as argued by Liebal, Behne, Carpenter, and Tomasello (2009), Tomasello et al. (2007), and van der Goot et al. (2014), among others; our argument is that the mere fact of distal pointing does not constitute evidence for a human-unique cognitive ability, given that apes also communicate triadically and distally in captivity (e.g., Leavens & Hopkins, 1998; Savage-Rumbaugh, Rumbaugh, & McDonald, 1985; the present study) and in the wild (Hobaiter et al., 2014; Veà & Sabater-Pi, 1998). A potentially fruitful future direction in this line of investigation (kindly suggested to us by an anonymous reviewer) would involve replications with apes of the protocols used by researchers with human infants to demonstrate infants' discrimination of different degrees of shared experience with others. For example, Liebal et al. (2009) demonstrated that 18-month-old human infants responded appropriately to experimenters' pointing gestures in accordance with the nature of the shared activities the children had had with those experimenters. It remains an open question whether great apes would similarly discriminate between people based on their different shared activities.

In conclusion, the present findings unambiguously refute claims to the effect that great apes do not communicate about entities from a distance. Either chimpanzees possess the same appreciation for a state of common ground with experimenters that van der Goot et al. (2014) claimed for their human 12-month-olds (i.e., this is a shared capacity of humans and chimpanzees), or—we think more likely—proximity of signaling does not index this appreciation.

References

Bakeman, R., & Quera, V. (2011). *Sequential analysis and observational methods for the behavioral sciences.* Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9781139017343

Bard, K. A. (1992). Intentional behavior and intentional communication in young free-ranging orangutans. *Child Development*, *62*, 1186–1197. doi:10.2307/1131526

Bard, K. A., Bakeman, R., Boysen, S. T., & Leavens, D. A. (2014). Emotional engagements predict and enhance social cognition in young chimpanzees. *Developmental Science*, *17*, 682–696. doi:10.1111/desc.12145

Bard, K. A., & Leavens, D. A. (2009). Socio-emotional factors in the development of joint attention in human and ape infants. In L. Röska-Hardy & E. M. Neumann-Held (Eds.), *Learning from animals? Examining the nature of human uniqueness* (pp. 89–104). London, UK: Psychology Press. doi:10.1891/1945-8959.10.1.9

Bard, K. A., & Leavens, D. A. (2014). The importance of development for comparative primatology. *Annual Review of Anthropology*, *43*, 183–200. doi:10.1146/annurev-anthro-102313-030223

Bates, E., Benigni, L., Bretherton, I., Camaioni, L., & Volterra, V. (1977). From gesture to the first word. In I. M. Lewis & L. Rosenblum (Eds.), *Interaction, conversation and the development of language* (pp. 247–308). New York, NY: Wiley.

Boesch, C. (2007). What makes us human *Homo sapiens*? The challenge of cognitive cross-species comparison. *Journal of Comparative Psychology, 121*, 227–240. doi:10.1037/0735-7036.121.3.227

Boesch, C. (2012). *Wild cultures: A comparison between chimpanzee and human cultures*. Cambridge, UK: Cambridge University Press. doi:10.1017/cbo9781139178532

Burkart, J. M., Hrdy, S. B., & van Schaik, C. P. (2009). Cooperative breeding and human cognitive evolution. *Evolutionary Anthropology, 18*, 175–186. doi:10.1002/evan.20222

Carpenter, M., & Call, J. (2013). How joint is the joint attention of apes and human infants? In J. Metcalfe & H. S. Terrace (Eds.), *Agency and joint attention* (pp. 49–61). New York, NY: Oxford University Press. doi:10.1093/acprof:oso/9780199988341.003.0003

Carpenter, M., Tomasello, M., & Savage-Rumbaugh, S. (1995). Joint attention and imitative learning in children, chimpanzees, and enculturated chimpanzees. *Social Development, 4*, 217–237. doi:10.1111/j.1467-9507.1995.tb00063.x

Custance, D. M., Whiten, A., & Bard, K. A. (1995). Can young chimpanzees (*Pan troglodytes*) imitate arbitrary actions? Hayes & Hayes (1952) revisited. *Behaviour, 132*, 837–859. doi:10.1163/156853995x00036

Desrochers, S., Morissette, P., & Ricard, M. (1995). Two perspectives on pointing in infancy. In C. Moore & P. J. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 85–101). Hillsdale, NJ: Erlbaum.

Fox, N. A., Henderson, H. A., Marshall, P. J., Nichols, K. E., & Ghera, M. M. (2005). Behavioral inhibition: Linking biology and behavior within a developmental framework. *Annual Review of Psychology, 56*, 235–262. doi:10.1146/annurev.psych.55.090902.141532

Franco, F., & Butterworth, G. (1996). Pointing and social awareness: Declaring and requesting in the second year. *Journal of Child Language, 23*, 307–336. doi:10.1017/s0305000900008813

Froese, T., & Leavens, D. A. (2014). The direct perception hypothesis: Perceiving the intention of another's action hinders its precise imitation. *Frontiers in Psychology, 5*(65), 1–15. doi:10.3389/fpsyg.2014.00065

Garcia Coll, C., Kagan, J., & Reznick, J. S. (1984). Behavioral inhibition in young children. *Child Development, 55*, 1005–1019. doi:10.2307/1130152

Herrmann, E., Call, J., Hernandez-Lloreda, M. V., Hare, B., & Tomasello, M. (2007). Humans have evolved specialized skills of social cognition: The cultural intelligence hypothesis. *Science, 317*, 1360–1366. doi:10.1126/science.1146282

Hobaiter, C., Leavens, D. A., & Byrne, R. W. (2014). Deictic gesturing in wild chimpanzees? Some possible cases. *Journal of Comparative Psychology, 128*, 82–87. doi:10.1037/a0033757

Hopkins, W. D., Russell, J. L., McIntyre, J., & Leavens, D. A. (2013). Are chimpanzees really so poor at understanding imperative pointing? Some new data and an alternative view of canine and ape social cognition. *PLoS ONE, 8*, e79338. doi:10.1371/journal.pone.0079338

Horowitz, A. C. (2003). Do humans ape? Or do apes human? Imitation and intention in humans (*Homo sapiens*) and other animals. *Journal of Comparative Psychology, 117*, 325–336. doi:10.1037/0735-7036.117.3.325

Hostetter, A. B., Russell, J. L., Freeman, H., & Hopkins, W. D. (2007). Now you see me, now you don't: Evidence that chimpanzees understand the role of the eyes in attention. *Animal Cognition, 10*, 55–62. doi:10.1007/s10071-006-0031-x

Hyatt, C. W., & Hopkins, W. D. (1998). Interspecies object exchange: Bartering in apes? *Behavioural Processes, 42*, 177–187. doi:10.1016/S0376-6357(97)00075-2

Kellogg, W. N., & Kellogg, L. A. (1933). *The ape and the child: A study of early environmental influence upon early behavior*. New York, NY: McGraw-Hill.

Leavens, D. A. (2004). Manual deixis in apes and humans. *Interaction Studies, 5*, 387–408. doi:10.1075/is.5.3.05lea

Leavens, D. A. (2012). Pointing: Contexts and instrumentality. In S. Pika & K. Liebal (Eds.), *Current developments in primate gesture research* (pp. 181–197). Amsterdam, Netherlands: Benjamins.

Leavens, D. A. (2014). The plight of the sense-making ape. In M. Cappuccio & T. Froese (Eds.), *Enactive cognition at the edge of sense-making* (pp. 81–104). Basingstoke, UK: Palgrave Macmillan. doi:10.1057/9781137363367.0010

Leavens, D. A., & Bard, K. A. (2011). Environmental influences on joint attention in great apes: Implications for human cognition. *Journal of Cognitive Education and Psychology, 10*, 9–31. doi:10.1891/1945-8959.10.1.9

Leavens, D. A., & Hopkins, W. D. (1998). Intentional communication by chimpanzees: A cross-sectional study of the use of referential gestures. *Developmental Psychology, 34*, 813–822. doi:10.1037//0012-1649.34.5.813

Leavens, D. A., & Hopkins, W. D. (1999). The whole hand point: The structure and function of pointing from a comparative perspective. *Journal of Comparative Psychology, 113*, 417–425. doi:10.1037//0735-7036.113.4.417

Leavens, D. A., & Hopkins, W. D. (2005). Multimodal concomitants of manual gesture by chimpanzees: Influence of food size and distance. *Gesture, 5*, 73–88. doi:10.1075/gest.5.1.07lea

Leavens, D. A., Hopkins, W. D., & Bard, K. A. (1996). Indexical and referential pointing in chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology, 110*, 346–353. doi:10.1037//0735-7036.110.4.346

Leavens, D. A., Hopkins, W. D., & Bard, K. A. (2005). Understanding the point of chimpanzee pointing: Epigenesis and ecological validity. *Current Directions in Psychological Science, 14*, 185–189. doi:10.1111/j.0963-7214.2005.00361.x

Leavens, D. A., Hopkins, W. D., & Bard, K. A. (2008). The heterochronic origins of explicit reference. In

J. Zlatev, T. Racine, C. Sinha, & E. Itkonen (Eds.), *The shared mind: Perspectives on intersubjectivity* (pp. 187–214). Amsterdam, Netherlands: Benjamins. doi:10.1075/celcr.12.12lea

Leavens, D. A., Hopkins, W. D., & Thomas, R. K. (2004). Referential communication by chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology, 118*, 48–57. doi:10.1037/0735-7036.118.1.48

Leavens, D. A., Racine, T. P., & Hopkins, W. D. (2009). The ontogeny and phylogeny of non-verbal deixis. In R. Botha & C. Knight (Eds.), *The prehistory of language* (pp. 142–165). Oxford, UK: Oxford University Press. doi:10.1093/acprof:oso/9780199545872.003.0008

Leavens, D. A., Russell, J. L., & Hopkins, W. D. (2005). Intentionality as measured in the persistence and elaboration of communication by chimpanzees (*Pan troglodytes*). *Child Development, 76*, 291–306. doi:10.1111/j.1467-8624.2005.00845.x

Lempers, J. D. (1979). Young children's production and comprehension of nonverbal deictic behaviors. *Journal of Genetic Psychology, 135*, 93–102. doi:10.1080/00221325.1979.10533420

Liebal, K., Behne, T., Carpenter, M., & Tomasello, M. (2009). Infants used shared experience to interpret pointing gestures. *Developmental Science, 12*, 264–271. doi:10.1111/j.1467-7687.2008.00758.x

Liszkowski, U., Schäfer, M., Carpenter, M., & Tomasello, M. (2009). Prelinguistic infants, but not chimpanzees, communicate about absent entities. *Psychological Science, 20*, 654–660. doi:10.1111/j.1467-9280.2009.02346.x

Lyn, H. (2010). Environment, methodology, and the object choice task in apes: Evidence for declarative comprehension and implications for the evolution of language. *Journal of Evolutionary Psychology, 8*, 333–349. doi:10.1556/jep.8.2010.4.3

Lyn, H., Russell, J. L., & Hopkins, W. D. (2010). The impact of environment on the comprehension of declarative communication in apes. *Psychological Science, 21*, 360–365. doi:10.1037/e598112013-060

Lyn, H., Russell, J. L., Leavens, D. A., Bard, K. A., Boysen, S. T., Schaeffer, J., & Hopkins, W. D. (2014). Apes communicate about absent and displaced objects: Methodology matters. *Animal Cognition, 17*, 85–94. doi:10.1007/s10071-013-0640-0

Mulcahy, N. J., & Hedge, V. (2012). Are great apes tested with an abject object-choice task? *Animal Behaviour, 83*, 313–321. doi:10.1016/j.anbehav.2011.11.019

Mundy, P., Kasari, C., & Sigman, M. (1992). Nonverbal communication, affective sharing, and intersubjectivity. *Infant Behavior and Development, 15*, 377–381. doi:10.1016/0163-6383(92)80006-g

Pedersen, J., Segerdahl, P., & Fields, W. M. (2009). Why apes point: Pointing gestures in spontaneous conversation of language-competent *Pan/Homo* bonobos. In E. Potocki & J. Krasinski (Eds.), *Primatology: Theories, methods, and research* (pp. 53–74). Hauppauge, NY: Nova.

Plooij, F. X. (1978). Some basic traits of language in wild chimpanzees? In A. Lock (Ed.), *Action, gesture and symbol* (pp. 111–131). London, UK: Academic Press.

Povinelli, D. J., Bierschwale, D. T., & Čech, C. G. (1999). Comprehension of seeing as a referential act in young children, but not juvenile chimpanzees. *British Journal of Developmental Psychology, 17*, 37–60. doi:10.1348/026151099165140

Povinelli, D. J., & Eddy, T. J. (1996). What young chimpanzees know about seeing. *Monographs of the Society for Research in Child Development, 61*(Serial No. 247). doi:10.2307/1166159

Racine, T. P., Leavens, D. A., Susswein, N., & Wereha, T. J. (2008). Conceptual and methodological issues in the investigation of primate intersubjectivity. In F. Morganti, A. Carassa, & G. Riva (Eds.), *Enacting intersubjectivity: A cognitive and social perspective to the study of interactions* (pp. 65–79). Amsterdam, Netherlands: IOS Press.

Reaux, J. E., Theall, L. A., & Povinelli, D. J. (1999). A longitudinal investigation of chimpanzees' understanding of visual perception. *Child Development, 70*, 275–290. doi:10.1111/1467-8624.00021

Russell, J. L., Lyn, H., Schaeffer, J. A., & Hopkins, W. D. (2011). The role of socio-communicative rearing environments in the development of social and physical cognition in apes. *Developmental Science, 14*, 1459–1470. doi:10.1111/j.1467-7687.2011.01090.x

Savage-Rumbaugh, E. S., Rumbaugh, D. M., & McDonald, K. (1985). Language learning in two species of apes. *Neuroscience & Biobehavioral Reviews, 9*, 653–665. doi:10.1016/0149-7634(85)90012-0

Thomas, E., Murphy, M., Pitt, R., Rivers, A., & Leavens, D. A. (2008). Understanding of visual attention by adult humans (*Homo sapiens*): A partial replication of Povinelli, Bierschwale, and Čech (1999). *Journal of Comparative Psychology, 122*, 428–436. doi:10.1037/0735-7036.122.4.428

Tomasello, M. (1995). Joint attention as social cognition. In C. Moore & P. J. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 103–130). Hillsdale, NJ: Erlbaum.

Tomasello, M. (2008). *Origins of human communication.* Cambridge, MA: MIT Press.

Tomasello, M., Call, J., & Gluckman, A. (1997). Comprehension of novel communicative signs by apes and human children. *Child Development, 68*, 1067–1080. doi:10.2307/1132292

Tomasello, M., Carpenter, M., & Liszkowski, U. (2007). A new look at infant pointing. *Child Development, 78*, 705–722. doi:10.1111/j.1467-8624.2007.01025.x

van der Goot, M. H., Tomasello, M., & Liszkowski, U. (2014). Differences in the nonverbal requests of Great Apes and human infants. *Child Development, 85*, 444–455. doi:10.1111/cdev.12141

Veà, J. J., & Sabater-Pi, J. (1998). Spontaneous pointing behaviour in the wild pygmy chimpanzee (*Pan paniscus*). *Folia Primatologica, 69*, 289–290. doi:10.1159/000021640

Volter, C. J., Rossano, F., & Call, J. (2015). From exploitation to cooperation: Social tool use in orang-utan mother-offspring dyads. *Animal Behaviour, 100*, 126–134. doi:10.1016/j.anbehav.2014.11.025

Whiten, A., Custance, D. M., Gómez, J.-C., Teixidor, P., & Bard, K. A. (1996). Imitative learning of artificial fruit processing in children (*Homo sapiens*) and chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology, 110*, 3–14. doi:10.1037//0735-7036.110.1.3

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's website:

Figure S1. Nahko, a 28-Year-Old Male Chimpanzee Gestures to Distant Bananas While Looking at the Cameraperson (Photograph by Lisa A. Reamer)

Figure S2. Chester, a 19-Year-Old Male Chimpanzee Points to Food While Looking at an Interlocutor (Photograph by Lisa A. Reamer)

Figure S3. Tanya, a 46-Year-Old Female Chimpanzee, Gestures Toward the Cameraperson From a Distance (Photograph by Emily Mocarski)

Figure S4. Tanya, a 46-Year-Old Female Chimpanzee, Gestures Toward a Person Holding Grapes From a Distance (Photograph by Mary Catherine Mareno)

Figure S5. Bischk a 29-Year-Old Female Chimpanzee Gestures From a Distance Toward a Person Holding Bananas (Photograph by Mary Catherine Mareno)

Child Development, September/October 2015, Volume 86, Number 5, Pages 1639–1652

Quality Counts: Developmental Shifts in Associations Between Romantic Relationship Qualities and Psychosocial Adjustment

Charlene Collibee and Wyndol Furman
University of Denver

The present study assessed a developmental task theory of romantic relationships by examining associations between romantic relationship qualities and adjustment across 9 years using a community-based sample of 100 male and 100 female participants (M_{age} Wave 1 = 15.83) in a Western U.S. city. Using multilevel modeling, the study examined the moderating effect of age on links between romantic relationship qualities and adjustment. Consistent with developmental task theory, high romantic quality was more negatively associated with internalizing symptoms and dating satisfaction during young adulthood than adolescence. Romantic relationship qualities were also associated with externalizing symptoms and substance use, but the degree of association was consistent across ages. The findings underscore the significance of romantic relationship qualities across development.

Throughout the lifespan, individuals encounter a series of developmental tasks that reflect widely held expectations regarding age-appropriate behavior (McCormick, Kuo, & Masten, 2011). One such developmental task is the emergence and maintenance of romantic relationships (Roisman, Masten, Coatsworth, & Tellegen, 2004). Developmental task theory proposes that romantic relationships are an emerging developmental task in adolescence and become a salient development task in adulthood (Roisman, Booth-LaForce, Cauffman, Spieker, & the NICHD Early Child Care Research Network 2009; Roisman et al., 2004). When it is an emerging developmental task, the quality of such romantic relationships is not expected to be especially related to future adaptation, because of the variability and instability of the transitional period. As an emerging task, adolescents may be less likely to benefit from higher quality relationships as they navigate new experiences and learn to manage new stressors associated with romantic relationships. However, the quality of romantic relationships once they are a salient developmental task is expected to be associated with future adaptation (Ford & Lerner, 1992; Roisman et al., 2004; Thelen & Smith, 1998). By the same reasoning, the quality of functioning within romantic relationships may be related to concurrent adjustment in adolescence and young adulthood, but such associations should be stronger in young adulthood. The present study tested this aspect of developmental task theory by examining developmental shifts in associations between romantic relationship qualities and adjustment from adolescence into young adulthood.

One tenant of developmental task theory is that romantic relationships change as a function of development (McCormick et al., 2011). Shifts in the nature of romantic relationships reflect the progression of romantic relationships from an emerging to a salient task. Indeed, such shifts also reflect the growing significance of these relationships for youth, such that as these experiences become more developed and salient, they should also become increasingly tied to adjustment. Consistent with this theory, as adolescents become older and transition into young adulthood, romantic partners increase in their importance and are more likely to serve as safe havens or secure bases (Hazan & Zeifman, 1994). Romantic relationships also become increasingly supportive, more intimate, and more interdependent as adolescents get older and move into young adulthood (Furman & Buhrmester, 1992; Giordano, Manning, Longmore, & Flanigan, 2012;

Preparation of this manuscript was supported by Grant 050106 from the National Institute of Mental Health (Wyndol Furman, P.I.) and Grant 049080 from the National Institute of Child Health and Human Development (Wyndol Furman, P.I.). Appreciation is expressed to the Project Star staff for their assistance in collecting the data, and to the Project Star participants and their partners, friends, and families.

Correspondence concerning this article should be addressed to Charlene Collibee, Department of Psychology, University of Denver, Frontier Hall, 2155 S. Race St., Denver, CO 80209. Electronic mail may sent to Charlene.Collibee@gmail.com.

© 2015 The Authors
Child Development © 2015 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2015/8605-0022
DOI: 10.1111/cdev.12403

Seiffge-Krenke, 2003). Furthermore, as adolescents get older, they begin to rank their romantic partners as higher on a hierarchy of support figures (Furman & Buhrmester, 1992). Romantic relationships also become more serious and committed; as adolescents move into young adulthood, they become increasingly likely to have relationships of longer duration and to report being in love (Furman & Winkles, 2011). Overall, the pattern suggests that adolescent romantic relationships may center around affiliation and sexual activity when they are first emerging, but they become increasingly significant and valued as adolescents transition to young adulthood and these experiences become a salient developmental task (Furman & Wehner, 1997).

The idea that romantic relationships not only become more common but also shift in their nature is a core tenant of developmental task theory. In addition to these developmental shifts, preliminary evidence suggests associations between quality of functioning within romantic relationships and adjustment exist. Indeed, a developmental task theory would suggest that the *quality* of functioning within salient developmental tasks may be most predictive of adaptation (Roisman et al., 2004). Among young adults, higher relationship quality is associated with more positive adjustment (Kamp Dush & Amato, 2005; Segrin, Powell, Givertz, & Brackin, 2003; Simon & Barrett, 2010). Yet, among adolescents, these associations are less consistent and not as well understood (cf. La Greca & Harrison, 2005; Zimmer-Gembeck, Siebenbruner, & Collins, 2001). Our limited understanding of these associations in adolescence reflects the fact that most studies of adolescents have examined whether one has a relationship or the degree of romantic involvement, rather than the quality of that involvement (Davila, 2008).

The associations between romantic involvement and adjustment have received significant attention, and the literature has consistently demonstrated links between romantic involvement and internalizing problems in adolescence, particularly among girls (Connolly & McIsaac, 2009; Davila, 2011; Joyner & Udry, 2000; Starr et al., 2012). However, the extant literature has not yet examined if and when the quality of romantic relationships emerge to serve protective functions for adjustment. More recently, prior work using the same data set as this study examined how associations between romantic involvement and adjustment change developmentally (Furman & Collibee, 2014). Specifically, having a romantic partner was associated with greater levels of substance use, externalizing symptoms, and internalizing symptoms in adolescence, but was associated with lower levels in young adulthood. However, this work did not examine the nature of that romantic involvement. An examination of the associations between romantic qualities and adjustment not only addresses a separate and significant gap in the literature, but also allows us to test a distinct and important aspect of a developmental task theory—that the quality of functioning within the task (i.e., romantic relationships) should become increasingly related to adjustment as the task becomes salient.

Additionally, most studies that consider the quality of the romantic experiences have only examined one age group. Studying the associations between the quality of romantic relationships and adjustment from adolescence to young adulthood longitudinally may be particularly illuminating in understanding the developmental course of romantic relationships as a developmental task. Specifically, we would anticipate that the quality of romantic relationships would become increasingly associated with adjustment as it becomes a salient developmental task. The few studies that have compared these links in adolescence and young adulthood offer evidence consistent with a developmental task theory. Specifically, romantic relationship security is more strongly associated with fewer externalizing symptoms in young adulthood than in adolescence (van Dulmen, Goncy, Haydon, & Collins, 2008). For individuals whose most significant peer is a romantic partner, greater commitment is associated with fewer emotional problems in early adulthood, but not in adolescence (Meeus, Branje, van der Valk, & de Wied, 2007). As such, this preliminary work indeed suggests that differential associations may exist between romantic relationship qualities and adjustment developmentally. However, these works each examine only two time points, limiting our understanding of the developmental course of these links, an important aspect of testing a developmental task theory. Additionally, only a few relationship qualities and indices of adjustment have been examined to date.

The current study assessed the associations between romantic relationship qualities and adjustment over a 9-year period from middle adolescence to young adulthood (15–24 years of age) as a test of developmental task theory. Specifically, we examined the links between romantic relationship quality (support, negative interactions, and romantic relationship satisfaction) and adjustment (internalizing symptoms, externalizing symptoms, substance use, and overall dating satisfaction), across time. We selected internalizing symptoms, externalizing symptoms, and substance use because they are broad and important aspects of adjustment and romantic relationships. We included relationship satisfaction as one of our indices of relationship quality, but we also included overall dating satisfaction as an index of adjustment because romantic relationship qualities likely influence feelings of dating satisfaction over time as well. Past research has examined satisfaction with a particular romantic relationship in adolescence (e.g., Levesque, 1993; Overbeek, Ha, Scholte, de Kemp, & Engels, 2007; Welsh, Haugen, Widman, Darling, & Grello, 2005), but we believe this is one of the first studies to examine *overall* satisfaction with dating life during adolescence. Overall dating satisfaction is an important construct that is somewhat distinct from satisfaction with a particular relationship, as it may be influenced not only by feelings toward a current partner, but also by the qualities of past relationships, the number of individuals they have seen romantically, and whether they have a current relationship or not.

We expected that the associations between romantic relationship qualities and adjustment would reflect that romantic relationships are an emerging developmental task in adolescence yet become a salient developmental task in young adulthood. Thus, we hypothesized that the associations between romantic quality and adjustment would vary as a function of age, such that we would see links between greater relationship quality and fewer adjustment problems across development, but that these links would get stronger as youth grow older and became young adults.

Method

Participants

The participants were part of a longitudinal study investigating the role of relationships with parents, peers, and romantic partners on psychosocial adjustment. Two hundred 10th-grade high school students (100 boys, 100 girls; M_{age} = 15 years 10.44 months old, *SD* = 0.49 years) were recruited. The participants came from 37 zip codes of working-class to upper-middle-class neighborhoods in a large Colorado metropolitan area. We sought to obtain a diverse sample by distributing brochures and sending letters to families residing in a number of different zip codes and to students enrolled in various schools in ethnically diverse neighborhoods. We were unable to determine the ascertainment rate because we used brochures and because the letters were sent to many families who did not have a 10th grader. We contacted interested families with the goal of selecting a sample that had an equal number of males and females, and had a distribution of racial/ethnic groups that approximated that of the United States. To ensure maximal response, we paid families $25 to hear a description of the project in their homes. Of the families that heard the description, 85.5% expressed interest and carried through with the Wave 1 assessment.

The sample consisted of 11.5% African Americans, 12.5% Hispanics, 1.5% Native Americans, 1% Asian American, 4% biracial, and 69.5% White, non-Hispanics. With regard to family structure, 57.5% were residing with two biological or adoptive parents, 11.5% were residing with a biological or adoptive parent and a stepparent or partner, and the remaining 31% were residing with a single parent or relative. The sample was of average intelligence (Wechsler Intelligence Scale for Children–III vocabulary score *M* = 9.8, *SD* = 2.44); 55.4% of their mothers had a college degree, indicating that the sample was predominately middle or upper middle class.

As part of the larger project, the participants and their mothers completed a number of other measures in Wave 1. Although these measures are not directly relevant to this particular study, we compared our sample's scores to comparable national norms of representative samples for trait anxiety scores on the State–Trait Anxiety Inventory (STAI; Spielberger, 1983), maternal report of externalizing symptoms on the Child Behavior Checklist (Achenbach, 1991), participants' reports of internalizing and externalizing symptoms on the Youth Self Report, and eight indices of substance use from the Monitoring the Future survey (Johnston, O'Malley, & Bachman, 2002). The present sample was more likely to have tried marijuana (54% vs. 40%, $z = 2.23$, $p < .05$); otherwise, the sample scores did not differ significantly from the national scores on the other 11 measures, including frequency of marijuana usage.

Procedure

Adolescents participated in a series of laboratory sessions in which they were interviewed about romantic relationships and completed questionnaires. The mother ($N = 169$) and a close friend ($N = 145$) nominated by the participant also completed questionnaires about the participant's psychosocial competence and risky/problem behaviors.

For the purposes of the current study, we used the first through seventh waves of data collection, beginning when the participants were in the 10th grade and ending approximately 5.5 years after graduation from high school. Data were collected on a yearly basis in Waves 1 through 4, and then 1½ years later for Waves 5–7. The seven waves of data were collected between 2000 and 2010. Participant retention was excellent (Waves 1 and 2: $N = 200$; Wave 3: $N = 199$; Wave 4: $N = 195$; Wave 5: $N = 186$; Wave 6: $N = 185$; Wave 7: $N = 179$). There were no differences on the variables of interest between those who did and did not remain in the study.

The study was approved by the local Institutional Review Board.

Measures

Network of Relationships Inventory: Behavioral Systems Version (Furman & Buhrmester, 2009)

Participants completed the Network of Relationships Inventory (NRI) to assess their perceptions of their most important romantic relationship in the last year. The short version of the NRI includes five items regarding social support and six items regarding negative interactions. Participants used a 5-point scale to rate how much each description was characteristic of their romantic relationship, and support and negative interaction scores were derived by averaging the relevant items (M_αs = .89 and .92, respectively).

Romantic Relationship Satisfaction

Relationship satisfaction was assessed through an adapted version of the Quality of Marriage Inventory (QMI; Norton, 1983). The QMI is a six-item self-report measure that assesses an individual's global perception of his or her relationship quality (Baxter & Bullis, 1986) and was adapted to assess relationship satisfaction among adolescents and young adults. An example of a question is, "My relationship with my boy/girlfriend makes me happy," which the participant then responds to on a 7-point Likert scale (1 = *strongly disagree/not at all true* to 7 = *strongly agree/very true*; $M_\alpha = .97$).

Dating Satisfaction

On the Dating History Questionnaire (Furman & Wehner, 1992) participants were asked, "How satisfied have you been with your romantic or dating life (or not dating, if you don't date)?" Satisfaction was rated on a 5-point scale. Convergent validity correlations can be found in Table 2.

Youth/Adult Self Report

Participants completed the Achenbach's (1991) Youth Self-Report in Waves 1–3 and Achenbach's (1997) Adult Self-Report in Waves 4–7. Internalizing and externalizing scores were derived from the 20 and 26 items that were comparable on the two versions (M_αs = .81 and .87, respectively).

Child/Adult Behavior Checklist

Friends and mothers reported on the participant's externalizing symptoms by completing the externalizing items of the Child Behavior Checklist in Waves 1–3 and the externalizing items on (1997) Adult Behavior Checklist in Waves 4–7 (Achenbach, 1991, 1997). Friend and mother reports of externalizing scores were derived from the 19 items that were comparable on the two versions (M_αs = .84 and .89, respectively).

Beck Depression Inventory

Participants completed the Beck Depression Inventory (BDI) to assess depressive symptoms (Beck, Rush, Shaw, & Emery, 1979; $M_\alpha = .86$).

State-Trait Anxiety Inventory

Participants completed the trait scale of Spielberger's (1983) STAI to assess anxious symptoms ($M_\alpha = .92$).

Drug Involvement Scale for Adolescence

Participants completed the Drug Involvement Scale for Adolescence (Eggert, Herting, & Thompson, 1996). For the present purposes, we examined use of beer, wine, liquor, marijuana, and other drugs (cocaine, opiate, depressants, tranquilizers, hallucinogens, inhalants, stimulants, over-the-counter drugs, club drugs) over the last 30 days. Frequency of each substance use was scored on a 7-point scale ranging from *never* to *every day*. Additionally, they completed a 16-item measure assessing adverse consequences arising from substance use ($M_\alpha = .94$) and an 8-item measure assessing difficulties in controlling substance use ($M_\alpha = .91$). The questionnaires on sexual behavior and substance use were administered by computer-assisted self-interviewing techniques to increase the candor of responses.

Friends Report of Substance Use

As part of their version of the Adolescent Self-Perception Profile (Harter, 1988), friends were asked five questions about the participant's use of alcohol and drugs and problems related to the use of those substances. The five items were averaged to derive the friend report of the participant's substance use and problems (M_α = .82).

Derivation of Composites

The participants', friends', and mothers' reports of externalizing symptoms were standardized across waves and averaged to derive an index of externalizing symptoms. BDI depression scores, STAI anxiety scores, and internalizing symptom scores were each standardized across waves and averaged to derive an index of internalizing symptoms. The various measures of the internalizing composite were correlated at M_r = .66 and externalizing measures were correlated at M_r = .32, and were all significant correlations.

The participants' reports of beer/wine drinking and their reports of drinking liquor were each standardized across waves and averaged to derive a measure of alcohol use. Similarly, the participants' reports of marijuana use and their reports of other drug use were standardized across waves and averaged to derive a measure of drug use. The participants' reports of problems and their reports of control problems were each standardized across waves and averaged to derive a measure of problem usage. The friends' reports of substance use were also standardized across waves. The alcohol, drug, and problem usage, and friends' reports of substance use were all significantly correlated (M_r = .55), and they were averaged to derive a measure of substance use.

Results

Preliminary Analyses

All variables were examined to ensure that they had acceptable levels of skew and kurtosis (Behrens, 1997). Outliers were winsorized to fall 1.5 times the interquartile range below the 25th percentile or above the 75th percentile. Additional descriptive statistics can be found in Table 1, and correlations among the study variables at Wave 1 and Wave 7 can be found in Table 2. Support and relationship satisfaction increased over time, though negative interactions did not (see Lantagne & Furman, 2015). In Wave 1, 59.8% of participants reported having had a romantic partner in the last year; in Wave 2, 66% had a romantic partner; in Wave 3, 78.2% had a romantic partner; in Wave 4, 75.9% had a romantic partner; in Wave 5, 73.5% had a romantic partner; in Wave 6, 79.9% had a romantic partner and in Wave 7, 80.6% had a romantic partner. Participant's reports of whether they had a romantic relationship or not was used as a measure of romantic involvement in the analyses.

Primary Analyses

We aimed to examine the pattern of associations between romantic quality and adjustment developmentally with a series of multilevel models using the statistical program Hierarchical Linear Modeling version 6.0, which uses full information maximum likelihood (FIML) to estimate parameters when data are missing (HLM; Raudenbush, Bryk, Cheong, & Congdon, 2001). FIML provides a powerful alternative to listwise deletion and protects against bias in analyses (Graham, Olchowski, & Gilreath, 2007; Little, Jorgensen, Lang, & Moore, 2014). When participants did not have a romantic relationship at a certain wave, romantic qualities

Table 1
Mean Romantic Involvement and Adjustment Outcomes (With Standard Deviations in Parentheses)

	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5	Wave 6	Wave 7
Age	15.88 (0.47)	16.89 (0.47)	17.94 (0.50)	19.03 (0.56)	20.51 (0.56)	22.11 (0.51)	23.70 (0.61)
Support	3.07 (1.03)	3.52 (1.08)	3.52 (1.05)	3.73 (1.00)	3.66 (1.04)	3.84 (0.98)	3.93 (0.96)
Negative interactions	1.82 (0.75)	1.71 (0.76)	1.95 (0.94)	1.74 (0.79)	1.88 (0.81)	1.79 (0.75)	1.74 (0.61)
Relationship satisfaction	11.34 (4.42)	12.28 (4.56)	12.03 (4.81)	13.63 (4.13)	12.26 (4.87)	13.10 (4.64)	13.18 (4.48)
Internalizing	0.16 (0.91)	0.16 (0.94)	0.05 (0.89)	−0.04 (0.86)	−0.04 (0.85)	−0.12 (0.83)	−0.18 (0.91)
Externalizing	0.27 (0.87)	0.13 (0.82)	0.14 (0.85)	0.04 (0.78)	−0.14 (0.66)	−0.24 (0.68)	−0.31 (0.68)
Substance use	−0.26 (0.49)	−0.14 (0.59)	−0.05 (0.63)	0.07 (0.58)	0.18 (0.65)	0.33 (0.61)	0.33 (0.65)
Dating satisfaction	3.25 (0.98)	3.38 (1.08)	3.40 (1.07)	3.56 (1.07)	3.54 (1.17)	3.61 (1.19)	3.64 (1.15)

Table 2
Patterns of Correlations Between Romantic Qualities and Adjustment Outcomes at (a) Wave 1 and (b) Wave 7

	Support	Negative interactions	Relationship satisfaction	Internalizing symptoms	Externalizing symptoms	Substance use	Dating satisfaction
(a) Support							
Negative interactions	−.07						
Relationship satisfaction	.64*	−.33*					
Internalizing symptoms	−.09	.08	−.04				
Externalizing symptoms	−.14	.21*	−.17	.40*			
Substance use	−.06	.06	−.20*	.04	.17*		
Dating satisfaction	.14	.06	.31*	−.14*	−.02	−.01	
Age	.07	.05	.02	.10	.03	−.01	.06
(b) Support							
Negative interactions	−.17						
Relationship satisfaction	.61*	−.36*					
Internalizing symptoms	−.27*	.35*	−.39*				
Externalizing symptoms	−.23*	.32*	−.20*	.49*			
Substance use	−.17	.16	−.23*	.15	.23*		
Dating satisfaction	.47*	−.21*	.50*	−.34*	−.30*	−.10	
Age	−.07	.05	.01	.00	.04	.07	−.04

*$p < .05$.

(i.e., support, negative interactions, and relationship satisfaction) were entered as missing. However, those participants who did not have a romantic relationship in a certain wave provided information on other variables of interest, including internalizing symptoms, externalizing symptoms, substance use, and romantic life satisfaction. Thus, existing data from all waves from all participants were used. To test our hypotheses regarding the developmental task theory, we used the following model.

Level 1:

$$\begin{aligned} Y_i = \beta_0 &+ \beta_1(\text{romantic involvement}) \\ &+ \beta_2(\text{age}) + \beta_3(\text{romantic quality}) \\ &+ \beta_4(\text{Age} \times \text{Romantic Quality}) + r_i \end{aligned}$$

Level 2:

$$\beta_0 = \gamma_{00} + \gamma_{01}(\text{gender}) + u_0$$

$$\beta_1 = \gamma_{10}$$

$$\beta_2 = \gamma_{20}$$

$$\beta_3 = \gamma_{30}$$

$$\beta_4 = \gamma_{40}$$

We used a two-step model to examine these associations, with both romantic qualities and age grand mean centered. First, we conducted a model with gender, romantic involvement, age, and romantic quality. We controlled for romantic involvement to assess if the quality of romantic relationships is associated with adjustment above the presence of a relationship. Next, we examined the interaction effects after the main effects to avoid concerns of conditionality (Little, 2013). We tested for significant random effects and found significant random slopes for age when predicting internalizing or externalizing symptoms using a $p < .20$ cutoff (Nezlek, 2012). In those instances, the equation included that random effect at Level 2. Table 3 reports the results of these analyses. Though all the effects are presented together, the unstandardized regression coefficients and standard errors are the values at the step in which these terms were first entered in the model. We also examined two-way and three-way interactions with gender, and found there were none.

Internalizing Symptoms

Romantic support. Consistent with a developmental task theory hypothesis, main effects of both age and romantic support on internalizing symptoms were found, but they were qualified by a significant interaction between age and romantic support (see Figure 1). To further interpret this and the other significant interactions, we used Preacher, Curran,

Table 3

Multilevel Models Testing the Associations Between Romantic Relationship Qualities and Adjustment

	Internalizing	Externalizing	Substance use	Dating satisfaction
Support				
Intercept (β_0)	0.00 (.05)	−0.02 (.04)	0.01 (.03)	3.58 (.13)
Romantic involvement (β_1)	0.04 (.04)	0.04 (.04)	0.57 (.23)*	0.33 (.48)
Age (β_2)	−0.05 (.01)**	−0.08 (.01)***	0.03 (.01)***	0.01 (.01)
Support (β_3)	−0.10 (.02)***	−0.06 (.02)**	−0.02 (.02)	0.41 (.03)***
Support × Age (β_4)	−0.02 (.01)*	−0.01 (.01)	−0.01 (.01)	0.03 (.01)*
Gender main effect (γ_{01})	0.23 (.10)*	−0.11 (.09)	−0.06 (.06)	−0.05 (.09)
Negative interactions				
Intercept (β_0)	0.06 (.09)	0.00 (.09)	−0.09 (.07)	3.51 (.05)
Romantic involvement (β_1)	−0.22 (.30)	0.06 (.28)	0.57* (.24)	0.51 (.51)
Age (β_2)	−0.05 (.01)***	−0.09 (.01)***	0.02 (.01)***	0.05 (.01)***
Negative Interactions (β_3)	0.21 (.03)*	0.13 (.03)***	0.05 (.02)*	−0.18 (.05)***
Negative interactions × Age (β_4)	0.02 (.01)*	0.01 (.01)	0.00 (.01)	−0.06 (.02)**
Gender main effect (γ_{01})	0.25 (.10)*	−0.11 (.09)	−0.08 (.06)	0.05 (.10)
Romantic relationship satisfaction				
Intercept (β_0)	−0.00 (.06)	0.01 (.05)	0.02 (.04)	3.56 (.07)
Romantic involvement (β_1)	0.01 (.15)	−0.01 (.12)	0.13 (.11)	0.26 (.22)
Age (β_2)	−0.05 (.01)***	−0.08 (.01)***	0.02 (.01)***	0.03 (.01)**
Relationship satisfaction (β_3)	−0.03 (.01)***	−0.02 (.004)**	−0.01* (.004)	0.10 (.01)***
Relationship Satisfaction × Age (β_4)	−0.01 (.00)**	−0.01 (.01)	−0.003 (.001)*	0.01 (.003)**
Gender main effect (γ_{01})	0.25 (.10)*	−0.13 (.09)	−0.08 (.06)	−0.01 (.08)

Note. The primary numbers in the table are the unstandardized coefficients for the fixed effects. Standard errors are in parentheses.
*$p < .05$. **$p < .01$. ***$p < .001$.

and Bauer's (2006) computational tools to plot the estimated effects of romantic qualities on change in the adjustment measure for two values of age: 1 *SD* below the mean age across all seven waves (16.66 years) and 1 *SD* above the mean age across all seven waves (21.92 years). In young adulthood, higher romantic support was associated with lower internalizing symptoms, $\beta = -.15$, $t(897) = -4.62$, $p < .001$. In adolescence, internalizing symptoms did not differ by degree of romantic support (see Figure 1).

Negative interactions. Consistent with hypotheses, a main effect of both age and negative interactions were found on internalizing symptoms, but again they were qualified by a significant interaction between age and negative interactions. As seen in Figure 2, higher negative interactions were associated with higher internalizing symptoms in both adolescence and young adulthood, but the association was stronger in young adulthood, $\beta = .16$, $t(902) = 3.95$, $p < .001$, and $\beta = .26$, $t(902) = 6.58$, $p < .001$, respectively.

Relationship satisfaction. Consistent with hypotheses, main effects of both age and romantic relationship satisfaction were found on internalizing symptoms, but again they were qualified by a significant interaction between age and relationship satisfaction. As seen in Figure 3, romantic relationship satisfaction was associated with lower internalizing symptoms in young adulthood, $\beta = -.06$,

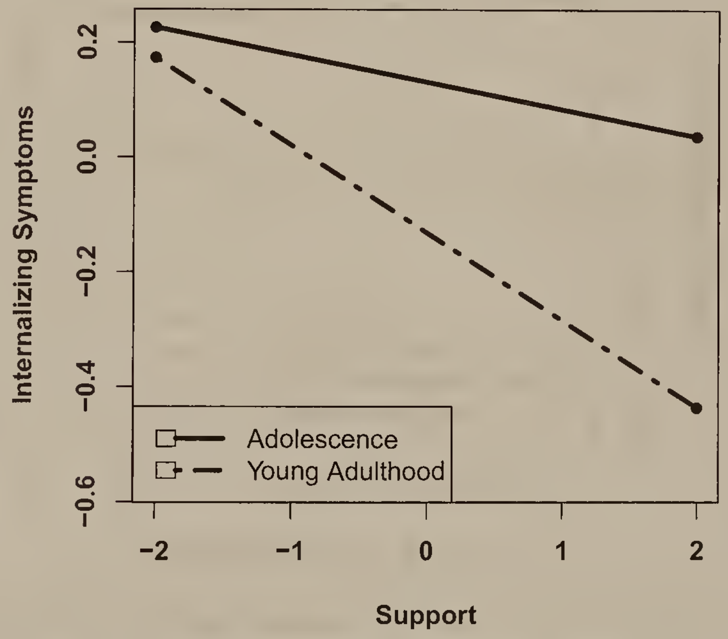


Figure 1. Interaction between support and age on internalizing symptoms. The two lines depict the association between support and internalizing symptoms at 1 *SD* below the mean of age (16.66 years, labeled adolescence) and 1 *SD* above the mean age (21.92 years, labeled young adulthood). Internalizing symptoms scores are standardized.

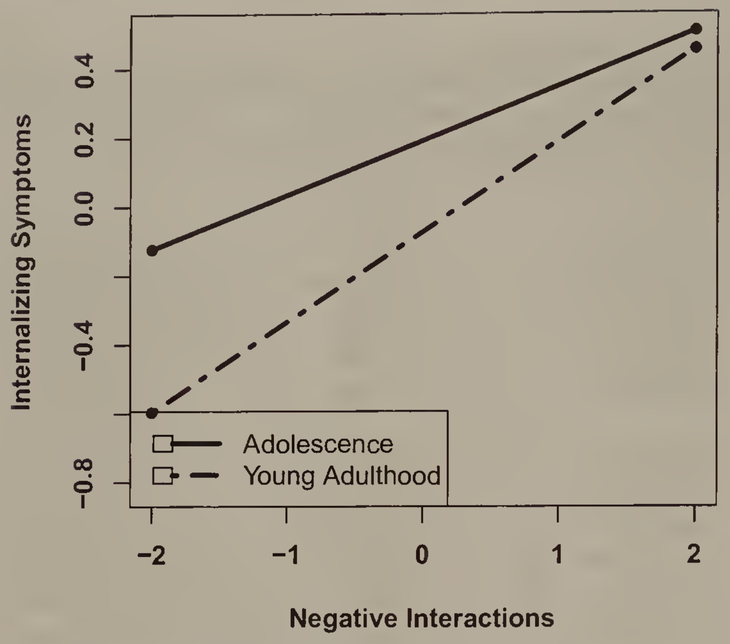


Figure 2. Interaction between negative interactions and age on internalizing symptoms. The two lines depict the association between negative interactions and internalizing symptoms at 1 *SD* below the mean of age (16.66 years, labeled adolescence) and 1 *SD* above the mean age (21.92 years, labeled young adulthood). Internalizing symptoms scores are standardized.

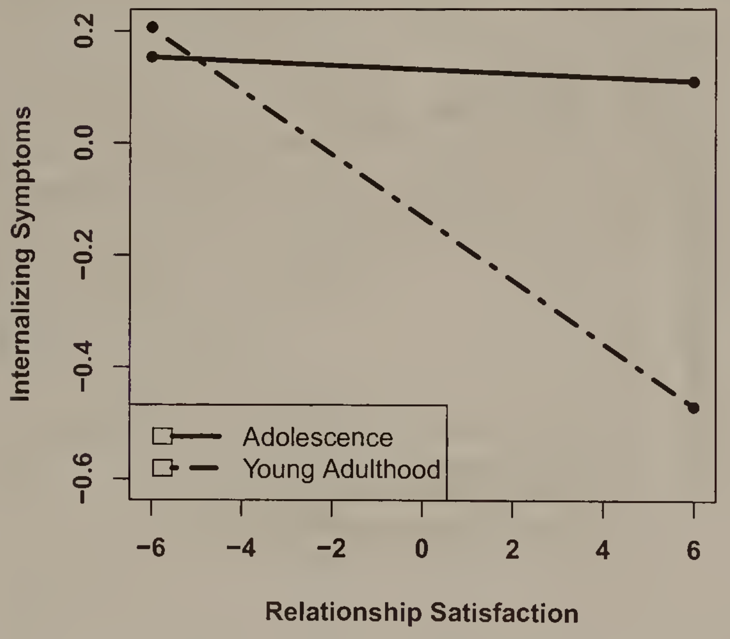


Figure 3. Interaction between relationship satisfaction and age on internalizing symptoms. The two lines depict the association between relationship satisfaction and internalizing symptoms at 1 *SD* below the mean of age (16.66 years, labeled adolescence) and 1 *SD* above the mean age (21.92 years, labeled young adulthood). Internalizing symptoms scores are standardized.

$t(898) = -5.45$, $p < .001$. In adolescence, internalizing symptoms did not differ by degree of relationship satisfaction.

Externalizing Symptoms

Romantic support. Main effects of both age and romantic support were found. There was a negative association with age, such that as individuals get older, they had fewer externalizing symptoms. Similarly, there was a negative association with romantic support, such that individuals with higher levels of romantic support had lower levels of externalizing symptoms. Contrary to hypotheses, no interaction between romantic support and age was found.

Negative interactions. Main effects of both age and negative interactions were found, such that as individuals got older they had fewer externalizing symptoms. Similarly, there was a positive association with negative interactions, such that individuals with higher levels of negative interactions had higher levels of externalizing symptoms. Contrary to hypotheses, no interaction between negative interactions and age was found.

Relationship satisfaction. Main effects of both age and romantic relationship satisfaction were found. There was a negative association with age, such that as individuals got older they had fewer externalizing symptoms. Similarly, there was a negative association with romantic relationship satisfaction, such that individuals with higher levels of relationship satisfaction had lower levels of externalizing symptoms. Contrary to hypotheses, no interaction between relationship satisfaction and age was found.

Substance Use

Romantic support. Regarding substance use, a main effect of age was found, such that as individuals got older they reported higher substance use. Contrary to hypotheses, neither the main effect of romantic support nor the interaction between romantic support and age was significant.

Negative interactions. Main effects of both age and negative interactions were found. There was a negative association with age, such that as individuals got older, they had fewer externalizing symptoms. Similarly, there was a positive association with negative interactions, such that individuals with higher levels of negative interactions had higher levels of substance use. Contrary to hypotheses, no interaction between negative interactions and age was found.

Relationship satisfaction. Regarding substance use, a main effect age was found, such that as individuals got older, they reported higher substance use.

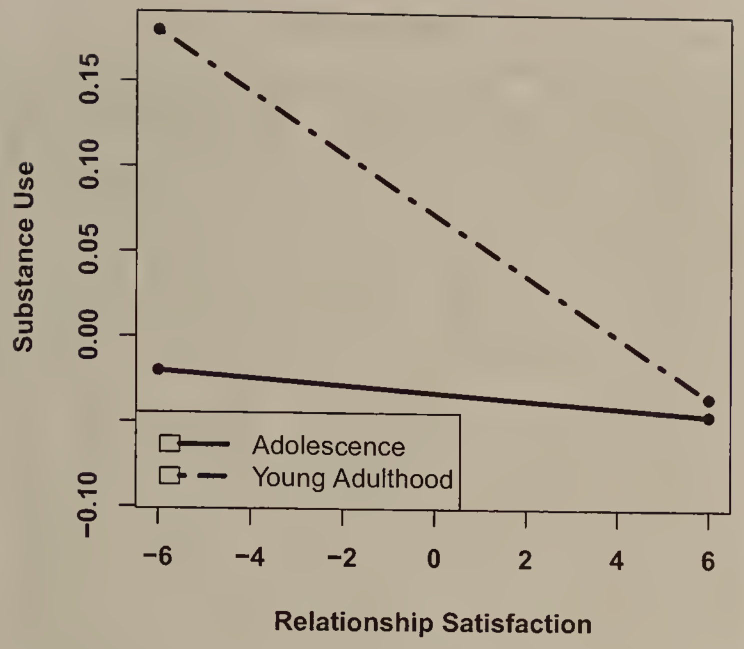


Figure 4. Interaction between relationship satisfaction and age on substance use. The two lines depict the association between relationship satisfaction and substance use at 1 *SD* below the mean of age (16.66 years, labeled adolescence) and 1 *SD* above the mean age (21.92 years, labeled young adulthood). Substance use scores are standardized.

There was no main effect of relationship satisfaction on substance use. Consistent with hypotheses, there was a significant interaction between age and relationship satisfaction on substance use. As seen in Figure 4, in young adulthood, high relationship satisfaction was associated with lower substance use, $\beta = -.02$, $t(905) = -3.74$, $p < .001$. In adolescence, substance use did not differ by degree of relationship satisfaction.

Dating Satisfaction

Romantic support. Consistent with hypotheses, a main effect of romantic support was found, but it was qualified by a significant interaction between age and romantic support. In both adolescence and young adulthood, high romantic support was associated with high dating satisfaction, but the association was stronger in young adulthood, $\beta = .33$, $t(884) = 8.30$, $p < .001$, and $\beta = .49$, $t(884) = 12.25$, $p < .001$, respectively (see Figure 5).

Negative interactions. Consistent with hypotheses, a main effect of negative interactions was found, but was qualified by a significant interaction between age and negative interactions on dating satisfaction. As seen in Figure 6, higher rates of negative interactions were associated with lower dating satisfaction in young adulthood, $\beta = -.34$, $t(889) = -4.66$, $p < .001$. In adolescence, dating satisfaction did not differ by degree of negative interactions.

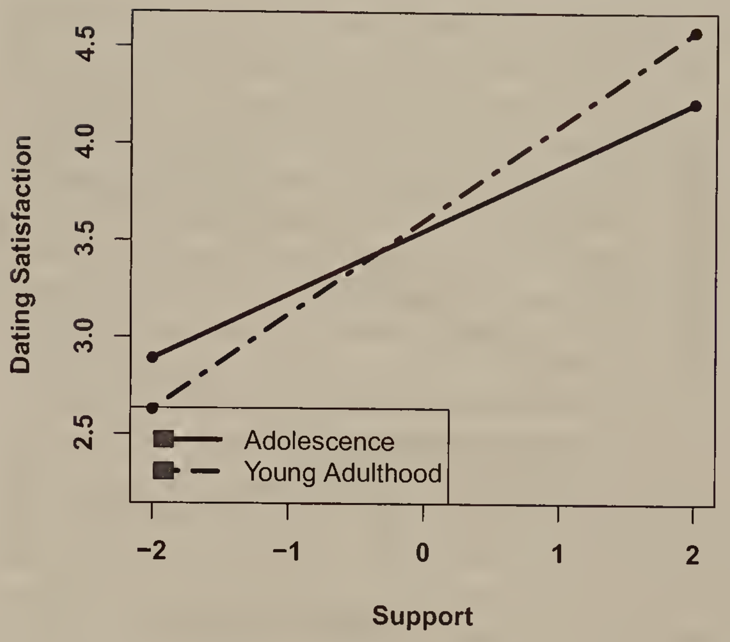


Figure 5. Interaction between support and age on dating satisfaction. The two lines depict the association between support and dating satisfaction at 1 *SD* below the mean of age (16.66 years, labeled adolescence) and 1 *SD* above the mean age (21.92 years, labeled young adulthood).

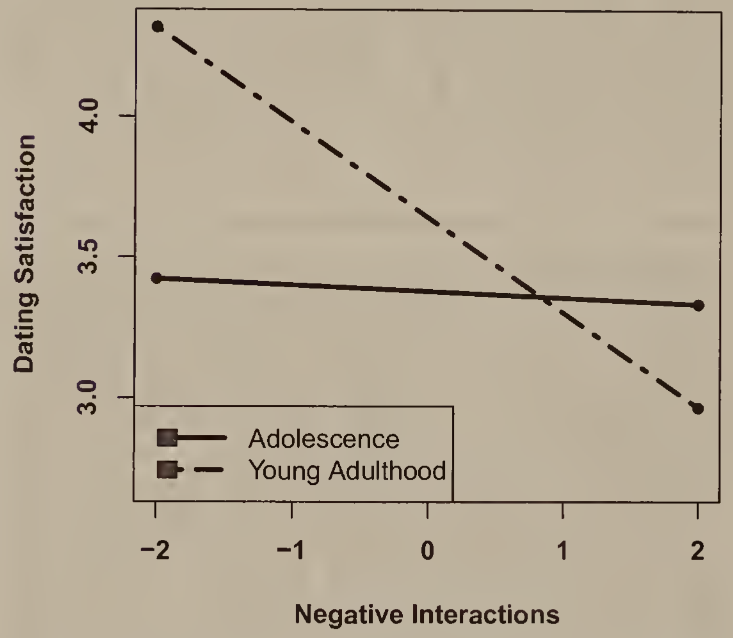


Figure 6. Interaction between negative interactions and age on dating satisfaction. The two lines depict the association between negative interactions and dating satisfaction at 1 *SD* below the mean of age (16.66 years, labeled adolescence) and 1 *SD* above the mean age (21.92 years, labeled young adulthood).

Relationship satisfaction. Consistent with hypotheses, main effects of relationship satisfaction and age were found, but were qualified by a significant interaction between age and relationship satisfaction on dating satisfaction. As seen in Figure 7,

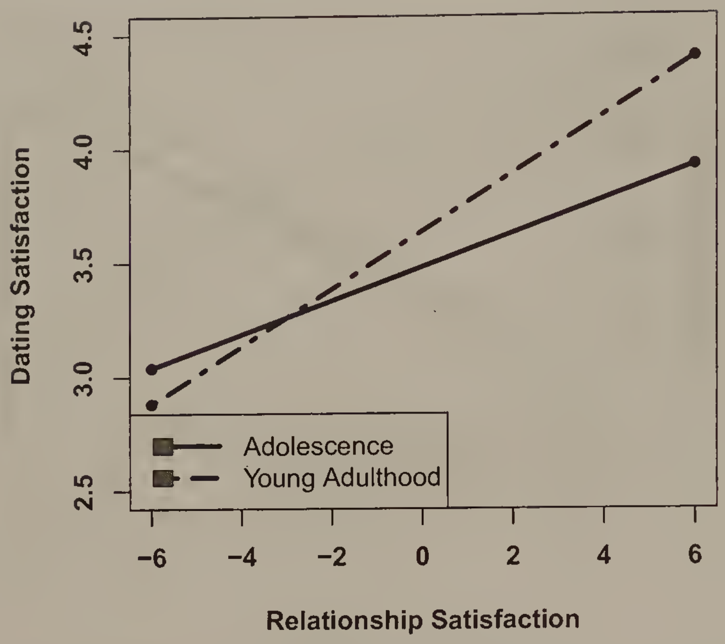


Figure 7. Interaction between relationship satisfaction and age on dating satisfaction. The two lines depict the association between relationship satisfaction and dating satisfaction at 1 *SD* below the mean of age (16.66 years, labeled adolescence) and 1 *SD* above the mean age (21.92 years, labeled young adulthood).

higher relationship satisfaction was associated with higher dating satisfaction in adolescence and young adulthood, though the association was greater in young adulthood, $\beta = .07$, $t(897) = 5.79$, $p < .001$, and $\beta = .13$, $t(897) = 9.92$, $p < .001$, respectively.

Secondary Analyses

The preceding analyses examine the links from romantic quality to adjustment. We next conducted a series of follow-up analyses examining the links from adjustment to romantic qualities, so as to better understand the potential directionality of these associations. We found significant main effects for internalizing symptoms and externalizing symptoms predicting lower support and relationship satisfaction, as well as greater negative interactions. Similarly, dating satisfaction predicted higher support and relationship satisfaction as well as lower rates of negative interactions. Substance use predicted greater negative interactions, but was not associated with support or relationship satisfaction. Regarding interactions between adjustment and age, the specific interactions varied, but the pattern of effects mirrored those in the primary analyses. In other words, in instances of significant interactions, adjustment became increasingly predictive of romantic qualities developmentally, consistent with developmental task theory. Results and plots are available as supplemental materials from the corresponding author.

Discussion

Past research has documented links between romantic relationship qualities and adjustment, but virtually all the work is cross-sectional in nature, examining one or at most two time points. The present study contributes to the literature by longitudinally examining the developmental course of associations between romantic relationship qualities and adjustment over a 9-year period spanning adolescence and young adulthood. Based on developmental task theory, we hypothesized that romantic relationship qualities would be associated with adjustment throughout this age period, but the quality of functioning within romantic relationships would become increasingly tied to concurrent adjustment as romantic relationships become a salient development task in young adulthood. As predicted, we found increasing associations between romantic relationship qualities and adjustment as youth got older in 7 of 12 instances; in 4 of the 5 remaining instances, significant main effects of romantic quality on adjustment were found. Notably, these main effects of quality were found even when controlling for relationship involvement.

Consistent with a developmental task theory, the associations between romantic relationship qualities and dating satisfaction were stronger in young adulthood than in adolescence. Such a pattern was observed for all three relationship qualities (support, negative interactions, and relationship satisfaction). These changing associations are particularly notable, as one might intuitively expect that romantic relationship qualities would be strongly associated with overall satisfaction with dating life at all ages. Indeed, prior work has demonstrated that adolescents prefer higher quality romantic partnerships (Levesque, 1993). However, these shifts can be accounted for by developmental task theory. During adolescence, when romantic relationships are an emerging developmental task, simply having a romantic relationship, having a number of relationships, or having a relationship with an attractive or otherwise desirable partner may play major roles in determining one's overall dating satisfaction, which may result in the quality of a relationship per se being less significant. In contrast, as romantic relationships become a salient developmental task, the quality of the relationship may become increasingly important to overall satisfaction. That is, the quality of a specific relationship becomes more predictive of overall dating satisfaction than other features of romantic life.

The findings regarding internalizing symptoms were also consistent with developmental task theory. As hypothesized, the associations between all three romantic relationship qualities and internalizing symptoms were stronger in young adulthood than in adolescence. When romantic relationships are a salient developmental task, poorer quality ones may lead to feelings of depression or anxiety. Alternatively, those who are depressed or anxious may have more difficulties establishing or maintaining high-quality relationships (Davila, 2008). Indeed, the supplemental analyses demonstrated significant associations from internalizing symptoms to romantic qualities. Thus, regardless of the directionality, it is clear that links between adjustment and romantic qualities are present across development and strengthen as romantic experiences become a salient task.

Finally, romantic relationship satisfaction was only associated with substance use in early adulthood. Perhaps, if substance use functions as a means of coping with dissatisfaction, then the increasing saliency of this dissatisfaction may be linked to higher levels of relying on substances to manage discontent. We did not, however, find similar interactions with age for either support or negative interactions. It is possible that support and negative interactions may not trigger such direct affective responses, which then may lead to maladaptive coping. Given the differences in the findings among the different qualities it will be important to replicate them before drawing firm conclusions.

The significant interactions with age for overall dating satisfaction, internalizing symptoms, and substance use are consistent with a developmental task theory of romantic relationship development. One explanation why romantic qualities become increasingly associated with adjustment is a stress and coping model of romantic experiences. This model posits that romantic experiences in adolescence may be stressful and overwhelming, thereby placing youth at risk for adjustment problems (Davila, 2008). In other words, adolescents may lack the necessary emotion regulation or coping skills to manage the stressors associated with romantic relationships. Though one might expect that romantic stressors and the quality of romantic relationships would be somewhat related, the challenges adolescents face in coping with most romantic relationships may be more related to simply initiating and maintaining the relationships, and thus the quality of a relationship may play a less significant role in affecting adjustment. Furthermore, the stressors associated with romantic involvement may be unrelated to the quality of the specific relationship. For example, adolescents may need to learn how to balance desires to spend time with a romantic partner and other academic or familial obligations. As adolescents transition into young adulthood, this model posits that they develop more coping skills and experience to better manage romantic stressors, and therefore may be able to benefit more from the social and emotional resources high-quality romantic relationships offer.

Contrary to our hypotheses, age did not moderate the associations between romantic relationship qualities and externalizing symptoms as well as most substance use variables. Substance use in particular has the greatest variability in patterns of findings. This variability may stem from the fact that the associations between romantic relationship quality and substance use depend on the patterns of partner substance use in young adulthood (Fleming, White, & Catalano, 2010). Further research examining these developmental changes should therefore incorporate assessments of partner substance use to better understand these associations.

Main effects of romantic quality were also found for externalizing symptoms such that greater support, greater relationship satisfaction, and fewer negative interactions were associated with fewer externalizing symptoms across adolescence and young adulthood. Indeed, other empirical evidence is consistent with these results. Specifically, romantic partner support is associated with less criminality in both adolescence and young adulthood (Meeus, Branje, & Overbeek, 2004).

What might account for the difference between the pattern of associations with externalizing and those with internalizing symptoms or dating satisfaction? One possibility is that externalizing symptoms are more stable than internalizing symptoms or dating satisfaction, and as a consequence, the externalizing symptoms' associations with other variables may be less likely to increase with age. Consistent with this idea, the correlation between externalizing symptoms in Wave 1 and Wave 7 was significantly greater than the corresponding correlation for internalizing symptoms ($r = .48$ vs. $r = .31$ using Fisher's r-to-z transformation, ZPF = 2.19, $p = .03$). Similarly, the correlation between externalizing symptoms in Wave 1 and Wave 7 was significantly greater than the corresponding correlation for dating satisfaction ($r = .48$ vs. $r = .22$, ZPF = 2.95, $p = .003$).

The differences in the pattern of results may also have occurred because different facets of romantic experiences are associated with different indices of adjustment. For example, the stress and coping model was derived as an explanation for internalizing symptoms and depression, not externalizing symptoms (Davila, 2008). Perhaps alternative mechanisms, such as partner characteristics are associated with externalizing behavior, such as delinquency. Adolescents who engage in delinquent behavior are likely to have delinquent partners (Lonardo, Giordano, Longmore, & Manning, 2009), and having such a romantic partner may lead to delinquent behavior (Haynie, Giordano, Manning, & Longmore, 2005).

The pattern of findings for externalizing symptoms is also distinct from prior work with this sample on romantic involvement (Furman & Collibee, 2014). In prior work, the presence of a romantic relationship was associated with more externalizing symptoms in adolescence, but less in young adulthood. In other words, the links between romantic involvement and externalizing symptoms vary by age, whereas the links between romantic relationship qualities and externalizing do not. Perhaps the negative links in adolescence are accounted for by the typical quality of these romantic relationships. As romantic relationships improve in quality developmentally, the associations between romantic involvement and adjustment may also change such that romantic involvement becomes protective in young adulthood.

Limitations and Future Directions

In the present study, we examined the longitudinal developmental course of associations between romantic relationship qualities and adjustment from adolescence into young adulthood. However, we only examined the concurrent associations at the different ages. Developmental task theory hypothesizes that once romantic relationships are a salient developmental task, quality of functioning in the task (i.e., romantic relationship qualities) should predict future adaptation (Roisman et al., 2004). Future research should examine such longitudinal associations. Studies examining whether romantic relationship qualities are predictive of success in other subsequent developmental tasks would be particularly informative.

Longitudinal studies would also help test potential alternative explanations for the present associations, as the concurrent nature of the present study does not address the directionality of the links. We focused on how romantic relationship qualities may be linked to adjustment, but the links may also go from adjustment to romantic relationship qualities, which would also be consistent with developmental task theory. Indeed, we examined and found the expected associations from adjustment to romantic qualities as well, suggesting that these links may go in either direction, which is consistent with some extant literature (Davila, Stroud, & Starr, 2009). In fact, several of our explanations of the associations also were based on the idea that adjustment may affect the quality of romantic relationships (e.g., internalizing symptoms affecting relationship qualities). Further work should aim to better understand these potential bidirectional processes. On a related note, the various relationship qualities were correlated with one another, as were the various adjustment indices. An important next step would be to better understand which romantic qualities may uniquely be related to specific adjustment indices.

Third-variable explanations such as partner effects may also be responsible for patterns of findings. Additionally, although the present study examined the developmental course of these links over almost a decade, expanding the age range being studied would be an important next step. For example, the links between romantic quality and adjustment may become even stronger as young adults begin cohabitating or get married, as links between marriage and well-being are well established (Williams, Frech, & Carlson, 2010).

Finally, we have demonstrated that the quality of romantic relationships is predictive of adjustment even when controlling for the presence of a romantic relationship. Future work should extend these findings and explore if the normative quality of romantic relationships in adolescence and young adulthood may partially account for the links between romantic involvement and adjustment in the literature.

Despite these limitations, the present study contributes significantly to the field by providing the first examination of the developmental course of links between romantic relationship qualities and adjustment over a 9-year span. Taken together, the present study provides evidence for a developmental task theory of romantic relationships. Perhaps more important, it offers consistent evidence of associations between romantic relationship qualities and adjustment across development. Indeed, higher romantic quality was commonly associated with adjustment in adolescence, and these links are magnified in young adulthood as romantic relationships become a salient developmental task.

References

Achenbach, T. M. (1991). *Child Behavior Checklist/4–18*. Burlington: University of Vermont.

Achenbach, T. M. (1997). *Manual for the Young Adult Self-Report and Young Adult Behavior Checklist*. Burlington: Department of Psychiatry, University of Vermont.

Baxter, L., & Bullis, C. (1986). Turning points in developing romantic relationships. *Human Communication Research, 12*, 469–493. doi:10.1111/j.1468-2958.1986.tb00088.x

Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, D. (1979). *Cognitive therapy of depression*. New York, NY: Guilford.

Behrens, J. T. (1997). Principles and procedures of exploratory data analysis. *Psychological Methods, 2*, 131–160. doi:10.1037/1082-989X.2.2.131

Connolly, J. A., & McIsaac, C. (2009). Romantic relationships in adolescence. In R. Lerner & L. Steinberg (Eds.), *Handbook of adolescent psychology* (3rd ed., pp. 104–151). Hoboken, NJ: Wiley.

Davila, J. (2008). Depressive symptoms and adolescent romance: Theory, research, and implications. *Child Development Perspectives, 2*, 26–31. doi:10.1111/j.1750-8606.2008.00037.x

Davila, J. (2011). Romantic relationships and mental health in emerging adulthood. In K. L. Fingerman & M. Cui (Eds.), *Romantic relationships in emerging adulthood* (pp. 275–292). New York, NY: Cambridge University Press.

Davila, J., Stroud, C. B., & Starr, L. R. (2009). Depression in couples and families. In I. Gotlib & C. Hammen (Eds.), *Handbook of depression* (2nd ed., pp. 467–491). New York, NY: Guilford.

Eggert, L. L., Herting, J. R., & Thompson, E. A. (1996). The drug involvement scale for adolescents (DISA). *Journal of Drug Education, 26*, 101–130. doi:10.2190/EQ6J-D4GH-K4YD-XRJB

Fleming, C. B., White, H. R., & Catalano, R. F. (2010). Romantic relationships and substance use in early adulthood: An examination of the influences of relationship type, partner substance use, and relationship quality. *Journal of Health and Social Behavior, 51*, 153–167. doi:10.1177/0022146510368930

Ford, D. H., & Lerner, R. M. (1992). *Developmental systems theory: An integrative approach*. Newbury Park, CA: Sage.

Furman, W., & Buhrmester, D. (1992). Age and sex differences in perceptions of networks of personal relationships. *Child Development, 63*, 103–115. doi:10.1111/j.1467-8624.1992.tb03599.x

Furman, W., & Buhrmester, D. (2009). Methods and measures: The network of relationships inventory: Behavioral systems version. *International Journal of Behavioral Development, 33*, 470–478.

Furman, W., & Collibee, C. (2014). A matter of timing: Developmental theories of romantic involvement and psychosocial adjustment. *Development and Psychopathology, 26*, 1149–1160. doi:10.1017/S0954579414000182

Furman, W., & Wehner, E. A. (1992). *Dating History Questionnaire*. Unpublished measure, University of Denver, Denver, CO.

Furman, W., & Wehner, E. A. (1997). Adolescent romantic relationships: A developmental perspective. *New Directions for Child and Adolescent Development, 1997*, 21–36. doi:10.1002/cd.23219977804

Furman, W., & Winkles, J. K. (2011). Transformations in heterosexual romantic relationships across the transition into adulthood: "Meet me at the bleachers . . . I mean the bar." In B. Laursen & W. A. Collins (Eds.), *Relationship pathways: From adolescence to young adulthood* (pp. 209–213). Thousand Oaks, CA: Sage.

Giordano, P. C., Manning, W. D., Longmore, M. A., & Flanigan, C. M. (2012). Developmental shifts in the character of romantic and sexual relationships from adolescence to young adulthood. In A. Booth, S. L. Brown, N. S. Landale, W. D. Manning, & S. M. McHale (Eds.), *Early adulthood in a family context* (pp. 133–164). New York, NY: Springer.

Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science, 8*, 206–213. doi:10.1007/s11121-007-0070-9

Harter, S. (1988). *Manual for the Self-Perception Profile for Adolescents*. Denver, CO: University of Denver.

Haynie, D. L., Giordano, P. C., Manning, W. D., & Longmore, M. A. (2005). Adolescent romantic relationships and delinquency involvement. *Criminology, 43*, 177–210. doi:10.1111/j.0011-1348.2005.00006.x

Hazan, C., & Zeifman, D. (1994). Sex and the psychological tether. In K. Bartholomew & D. Perlman (Eds.), *Attachment processes in adulthood: Advances in personal relationships* (pp. 151–178). London, UK: Kingsley.

Johnston, L. D., O'Malley, P. M., & Bachman, J. G. (2002). *Monitoring the Future national results on adolescent drug use: Overview of key findings, 2001* (NIH Publication No. 02-5105). Bethesda, MD: National Institute on Drug Abuse.

Joyner, K., & Udry, J. R. (2000). You don't bring me anything but down: Adolescent romance and depression. *Journal of Health and Social Behavior, 41*, 369–391. doi:10.2307/2676292

Kamp Dush, C. M., & Amato, P. R. (2005). Consequences of relationship status and quality for subjective well-being. *Journal of Social and Personal Relationships, 22*, 607–627. doi:10.1177/0265407505056438

La Greca, A. M., & Harrison, H. M. (2005). Adolescent peer relations, friendships, and romantic relationships: Do they predict social anxiety and depression? *Journal of Clinical Child and Adolescent Psychology, 34*, 49–61. doi:10.1207/s15374424jccp3401_5

Lantagne, A. S., & Furman, W. (2015). *Emerging qualities of romantic relationships from adolescence to adulthood*. Unpublished manuscript.

Levesque, R. J. (1993). The romantic experience of adolescents in satisfying love relationships. *Journal of Youth and Adolescence, 22*, 219–251. doi:10.1007/BF01537790

Little, T. D. (2013). *Longitudinal structural equation modeling*. New York, NY: Guilford.

Little, T. D., Jorgensen, T. D., Lang, K. M., & Moore, E. W. G. (2014). On the joys of missing data. *Journal of Pediatric Psychology*, *39*, 151–162. doi:10.1093/jpepsy/jst048

Lonardo, R. A., Giordano, P. C., Longmore, M. A., & Manning, W. D. (2009). Parents, friends, and romantic partners: Enmeshment in deviant networks and adolescent delinquency involvement. *Journal of Youth and Adolescence*, *38*, 367–383. doi:10.1007/s10964-008-9333-4

McCormick, C. M., Kuo, S. I. C., & Masten, A. S. (2011). Developmental tasks across the lifespan. In K. L. Fingerman, C. A. Berg, J. Smith, & T. C. Antonucci (Eds.), *Handbook of lifespan development* (pp. 117–140). New York, NY: Springer.

Meeus, W., Branje, S., & Overbeek, G. J. (2004). Parents and partners in crime: A six-year longitudinal study on changes in supportive relationships and delinquency in adolescence and young adulthood. *Journal of Child Psychology and Psychiatry*, *45*, 1288–1298. doi:10.1111/j.1469-7610.2004.00312.x

Meeus, W. H., Branje, S. J., van der Valk, I., & de Wied, M. (2007). Relationships with intimate partner, best friend, and parents in adolescence and early adulthood: A study of the saliency of the intimate partnership. *International Journal of Behavioral Development*, *31*, 569–580. doi:10.1177/0165025407080584

Nezlek, J. B. (2012). Multilevel modeling analysis of diary-style data. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 357–383). New York, NY: Guilford.

Norton, R. (1983). Measuring marital quality: A critical look at the dependent variable. *Journal of Marriage and the Family*, *45*, 141–151. doi:10.2307/351302

Overbeek, G., Ha, T., Scholte, R., de Kemp, R., & Engels, R. C. (2007). Brief report: Intimacy, passion, and commitment in romantic relationships—Validation of a "triangular love scale" for adolescents. *Journal of Adolescence*, *30*, 523–528. doi:10.1016/j.adolescence.2006.12.002

Preacher, K. J., Curran, P. J., & Bauer, D. J. (2006). Computational tools for probing interactions in multiple linear regression, multilevel modeling, and latent curve analysis. *Journal of Educational and Behavioral Statistics*, *31*, 437–448. doi:10.3102/10769986031004437

Raudenbush, S., Bryk, A., Cheong, Y. F., & Congdon, R. (2001). *HLM5: Hierarchical linear and nonlinear modeling*. Chicago, IL: Scientific Software International.

Roisman, G. I., Booth-LaForce, C., Cauffman, E., Spieker, S., & the NICHD Early Child Care Research Network. (2009). The developmental significance of adolescent romantic relationships: Parent and peer predictors of engagement and quality at age 15. *Journal of Youth and Adolescence*, *38*, 1294–1303. doi:10.1007/s10964-008-9378-4

Roisman, G. I., Masten, A. S., Coatsworth, J. D., & Tellegen, A. (2004). Salient and emerging developmental tasks in the transition to adulthood. *Child Development*, *75*, 123–133. doi:10.1111/j.1467-8624.2004.00658.x

Segrin, C., Powell, H. L., Givertz, M., & Brackin, A. (2003). Symptoms of depression, relational quality, and loneliness in dating relationships. *Personal Relationships*, *10*, 25–36. doi:10.1111/1475-6811.00034

Seiffge-Krenke, I. (2003). Testing theories of romantic development from adolescence to young adulthood: Evidence of a developmental sequence. *International Journal of Behavioral Development*, *27*, 519–531. doi:10.1080/01650250344000145

Simon, R. W., & Barrett, A. E. (2010). Nonmarital romantic relationships and mental health in early adulthood: Does the association differ for women and men? *Journal of Health and Social Behavior*, *51*, 168–182. doi:10.1177/0022146510372343

Spielberger, C. D. (1983). *Manual for the State-Trait Anxiety Inventory—STAI*. Palo Alto, CA: Consulting Psychologists Press.

Starr, L. R., Davila, J., Stroud, C. B., Li, P. C. C., Yoneda, A., Hershenberg, R., & Miller, M. R. (2012). Love hurts (in more ways than one): Specificity of psychological symptoms as predictors and consequences of romantic activity among early adolescent girls. *Journal of Clinical Psychology*, *68*, 403–420. doi:10.1002/jclp.20862

Thelen, E., & Smith, L. B. (1998). Dynamic systems theories. In R. M. Lerner (Vol. Ed.) & W. Damon & R. M. Lerner (Series Eds.), *Handbook of child psychology: Vol. 1. Theoretical models of human development:* pp. 258–312). Hoboken, NJ: Wiley.

van Dulmen, M. H., Goncy, E. A., Haydon, K. C., & Collins, W. A. (2008). Distinctiveness of adolescent and emerging adult romantic relationship features in predicting externalizing behavior problems. *Journal of Youth and Adolescence*, *37*, 336–345. doi:10.1007/s10964-007-9245-8

Welsh, D. P., Haugen, P. T., Widman, L., Darling, N., & Grello, C. M. (2005). Kissing is good: A developmental investigation of sexuality in adolescent romantic couples. *Sexuality Research & Social Policy*, *2*, 32–41. doi:10.1525/srsp.2005.2.4.32

Williams, K., Frech, A., & Carlson, D. L. (2010). Marital status and mental health. In T. L. Scheid & T. N. Brown (Eds.), *A handbook for the study of mental health: Social contexts, theories, and systems* (pp. 306–320). New York, NY: Cambridge University Press.

Zimmer-Gembeck, M. J., Siebenbruner, J., & Collins, W. A. (2001). Diverse aspects of dating: Associations with psychosocial functioning from early to middle adolescence. *Journal of Adolescence*, *24*, 313–336. doi:10.1006/jado.2001.0410

Child Development, September/October 2015, Volume 86, Number 5, Pages 1653–1671

Elementary School Children's Reasoning About Social Class: A Mixed-Methods Study

Rashmita S. Mistry
University of California–Los Angeles

Christia S. Brown
University of Kentucky

Elizabeth S. White
Illinois State University

Kirby A. Chow
Society for Research in Child Development/American Association for the Advancement of Science Policy Fellow

Cari Gillen-O'Neel
Macalester College

The current study examined children's identification and reasoning about their subjective social status (SSS), their beliefs about social class groups (i.e., the poor, middle class, and rich), and the associations between the two. Study participants were 117 10- to 12-year-old children of diverse racial, ethnic, and socioeconomic backgrounds attending a laboratory elementary school in Southern California. Results indicated that children's SSS ratings correlated with indicators of family socioeconomic status and were informed by material possessions, lifestyle characteristics, and social and societal comparisons. Children rated the poor as having fewer positive attributes and more negative attributes than the middle class, and fewer positive attributes than the rich. Lower SSS children held less positive attitudes toward the poor than children with middle SSS ratings.

Social class is both important and complex. Beyond simply labeling a person as poor, middle class, or rich, social class is:

> A context rooted in *both* the material substance of social life (wealth, education, work) *and* the individual's construal of his or her class rank, and is a core aspect of how he or she thinks of the self and relates to the social world. (emphasis added; Kraus, Piff, Mendoza-Denton, Rheinschmidt, & Keltner, 2012, p. 546)

Based on this conceptualization, social class consists of socioeconomic status (SES)—the actual household income, education level, and occupational prestige of a family—and subjective social status (SSS), or an individual's perception of their social standing or "place" in relation to others in society (American Psychological Association, Task Force on Socioeconomic Status, 2007; Diemer, Mistry, Wadsworth, López, & Reimers, 2013). Research with adults has repeatedly shown that SSS is as strong as, and sometimes an even stronger than, predictor of physical health outcomes among adults compared to more objective indicators of social class (Adler, Epel, Castellazzo, & Ickovics, 2000; Goodman et al., 2001; Singh-Manoux, Adler, & Marmot, 2003; Singh-Manoux, Marmot, & Adler, 2005). Adults' SSS is also associated with their thoughts, feelings, and behaviors toward others (see review by Kraus et al., 2012). Although extensive research has examined the association between SES and children's outcomes, little work has examined children's SSS or their understanding of social class. That is the focus of the current study.

Children's Understanding of Social Class: A Developmental Intergroup Perspective

Much of the developmental research on the social class of children examines the effects of

---

Correspondence concerning this article should be addressed to Rashmita S. Mistry, Department of Education, University of California Los Angeles, Los Angeles, CA 90095. Electronic mail may be sent to mistry@gseis.ucla.edu.

© 2015 The Authors
Child Development © 2015 Society for Research in Child Development, Inc.
All rights reserved. 0009-3920/2015/8605-0023
DOI: 10.1111/cdev.12407

poverty on developmental outcomes. The deleterious effects of poverty on mental and physical health, academic achievement, and a wide variety of outcomes are extremely important but beyond the scope of this article (see reviews by Aber, Morris, & Raver, 2012; McLoyd, Aikens, & Burton, 2006). Rather, our focus is on children's knowledge of and reasoning about social class: What do children think it means to be poor, middle class, or rich? And, importantly, which group do children think they belong to and why? On these topics, there is considerably less research.

The limited evidence that does exist intimates that children's perception of their own social class standing has consequences for their academic competencies, psychological well-being, and future career choices (e.g., Crosnoe, 2009; Mistry, Benner, Tan, & Kim, 2009; Weinger, 1998). In popular media, there are vivid and compelling accounts of what poverty looks like from a child's perspective (e.g., see Kotlowitz, 1992; Kozol, 1992). From these accounts, we know that children are often acutely aware of their families' precarious economic position. Within the scientific literature, qualitative studies have documented that children from low-income families are not only aware of the negative societal attitudes about those receiving welfare and food stamps, but feel embarrassed by their own families' use of such benefits and report being teased and ridiculed by peers because of it (Trzcinski, 2002). Furthermore, awareness of one's family's financial hardships is related to higher levels of depressive symptoms and lower levels of academic engagement among adolescents (Mistry et al., 2009). We do not know, however, how children conceptualize and reason about their own and others' social class, or what information (e.g., parents' income level) guides these conceptualizations.

To examine children's understanding of social class, we take a developmental intergroup theoretical framework (Bigler & Liben, 2006). Social classes (such as poor, middle class, and rich) are social groups, in many ways similar to racial, ethnic, and gender groups. These types of social groups are exclusive (e.g., an individual can only belong to one gender group), are marked by distinct characteristics, and have prototypic members. Thus, the ways in which children understand and develop attitudes about other social groups will likely be similar to how they develop attitudes about social class groups. As such, an intergroup framework (the broad term used to refer to children's understanding of and attitudes about social groups) is informative for guiding our own research. According to developmental intergroup theory (DIT; Bigler & Liben, 2006), children (and adults) naturally want to sort and categorize people into social groups. Children attend to those social groups that are perceptually and psychologically salient, and they then attach essentialist meaning to those categories or groups (see also Aboud, 1988; Ruble et al., 2004). In other words, children assume that the groups—if they are perceptually and psychologically salient—differ in important, innate ways from one another. They associate these salient groups with specific attributes—associations and beliefs derived from environmental messages and experiences (Bigler & Liben, 2006). In the current study, using a developmental intergroup framework, we examine children's conceptions of social class by: (a) assessing children's identification of their own social class group (i.e., SSS), (b) examining their associations and beliefs about social class groups (i.e., the poor, middle class, and rich), and (c) examining the relations between children's SSS and their beliefs about the poor, middle class, and rich (i.e., how their own perceived social class group identification is related to their beliefs about other social class groups).

We examine these issues among children within the upper elementary school years (ages 10–12), as theoretical and empirical evidence suggests that the upper elementary school years are an important time to investigate children's developing sense of self and beliefs about others based on social group membership (Bennett & Sani, 2004; Ruble et al., 2004). Children in this age group have the requisite cognitive skills to classify individuals into distinct social groups and can understand that some social groups are associated with greater social status than other groups (Bigler, Averhart, & Liben, 2003). Children in this age group also have developed a more complex moral reasoning than younger children, referred to as equity-based reasoning (Damon, 1988). This is important in the study of children's understanding of social class. Children in lower elementary school use primarily equality-based reasoning, in which fairness is dependent on everyone getting the exact same treatment. In contrast, children in upper elementary school better understand the concept of equity, which would allow them to hold the belief that some individuals may have more income than other individuals because of such factors as education, effort, or ability. Although they may not endorse these beliefs, it is important to examine these concepts in children who are cognitively capable of understanding equity (or the lack of equity).

Children's Social Class Group Identification and SSS

Within the field of developmental intergroup research, considerable work has examined *social group identification*, which refers to the ability to place oneself and others accurately and consistently within one or more social groups (Ruble et al., 2004; Tajfel, 1982). Gender, racial, and ethnic groups are typically considered the two more relevant and important social categories to children (e.g., DuBois, Burk-Braxton, Swenson, Tevendale, & Hardesty, 2002; Ruble et al., 2004), yet they are not the only important social groups for children.

Like race, ethnicity, and gender, children also experience the world (and how others view them) as members of a distinct social class, with the stereotypes and biases associated with lower versus higher social class groups (Weinger, 2000). Social class, however, is a complex social category that can be somewhat difficult to identify in the self and others (Diemer et al., 2013). For example, in comparison to gender, race, and ethnicity, there are fewer clear-cut, perceptually salient characteristics (e.g., hair length for gender or skin color for race) for identifying one's social class group. Additionally, social class cues can be manipulated, distorted, and even masked entirely, further obscuring one's actual group membership. For example, individuals may choose to downplay their socioeconomic position by maintaining a more modest lifestyle, whereas other individuals may choose a lifestyle that conveys greater social status yet stretches them financially. Despite this, for children, there may still be perceptually and psychologically salient markers of social class, such that some children may dress in more expensive clothes than others, discuss taking expensive vacations with their families, or have more expensive possessions (e.g., iPhones) than others. These material cues may serve as de facto markers of social class.

Indeed, evidence suggests that children are capable of distinguishing between individuals from different social class backgrounds (specifically rich and poor) from an early age. Ramsey (1991), for example, showed that preschoolers were able to sort objects into "rich" and "poor" categories. As would be suggested by DIT (Bigler & Liben, 2006), during the elementary school years, children focus on external or peripheral cues when distinguishing between wealth and poverty (e.g. Chafel, 1997; Chafel & Neitzel, 2005; Jahoda, 1959; Leahy, 1981, 1983). Children are capable of distinguishing possessions on the basis of economic status cues (Chafel, 1997; Emler & Dickinson, 1985; Jahoda, 1959; Leahy, 1981, 1983; Tudor, 1971; Weinger, 2000), and they show awareness of occupational status differences (e.g., Bigler et al., 2003; Cook et al., 1996; Simmons & Rosenberg, 1971; Weinger, 1998).

In the current study, our first research question assessed how children identified their *own* social class group. Very little extant research with children could inform hypotheses for this question. Among a small adolescent convenience sample of working- and upper-middle-class youth, Goodman et al. (2000) found that adolescents agree that family income, parental occupation, and family lifestyle are important determinants of social class. Consistent with this information, upper-middle-class adolescents were reasonably accurate at evaluating their own social class group. In contrast, most working-class youth subjectively identified as either middle class or upper middle class, not as working class, and aspired to be upper middle class as adults (Goodman et al., 2000). This pattern of identification among working-class youth is consistent with findings from adults—research with adults (both in America and internationally) shows an overwhelming tendency to subjectively identify as middle class, regardless of actual levels of income and wealth (Diemer & Ali, 2009; Evans & Kelley, 2004; Ladd & Bowman, 1998).

To our knowledge, the Goodman et al. (2000) study is one of the only quantitative studies examining adolescents' SSS, and no known studies have examined preadolescent children's SSS in a similar manner. Prior studies rely on parents' reports of family SES or social class background as inferred by characteristics of the child's neighborhood or school (e.g., percentage of students eligible for or receiving free or reduced school meals). Thus, one of the goals of the current study was to build upon the findings of Goodman et al. by examining the extent to which upper elementary school-age children are able to identify their own social class group and whether or not their identification aligns with more objective indicators of SES (i.e., family income, parent education level) and parents' SSS ratings. DIT (Bigler & Liben, 2006) argues that to be meaningful to children, social groups must be marked with perceptually and psychologically salient characteristics. Because researchers are less clear on which cues children use to determine social class group membership (compared to, for example, race), we were also interested in what information children use to evaluate their own social class group.

Children's Beliefs About Social Class Groups

Although researchers have not examined whether children in elementary school can identify their own social class group, some existing research has examined children's beliefs about the differences between social classes. Toward the end of elementary school (i.e., between the ages of 10 and 12), children begin to show links between internal attributes (e.g., working hard, being smart) and social class group membership (Leahy, 1981; Sigelman, 2012). Most of these beliefs are stereotypes in which being poor is associated with more negative attributes than being rich. For example, Woods, Kurtz-Costes, and Rowley (2004) observed that elementary and middle school students rated the rich as more academically competent than the poor. Skafte (1989) found that adolescents' ratings of strangers were significantly lower if the stranger was depicted as being poor than rich, and adolescents assigned more negative stereotypical attributes (e.g., stealing, making friends less easily) and fewer positive stereotypical attributes (e.g., more intelligent, having better grades, happier) to poor as compared to rich strangers. The findings are consistent with research on adult stereotypes about the poor—among adults, negative stereotypes about the poor as lazy, unwilling to work, not well educated, dishonest, and lacking in self-control are pervasive (Bullock, 1995; Cook & Curtin, 1987; Gans, 1995; Katz, 1989; Williams, 2009). Notably absent from the extant research, however, is children's beliefs about the middle class. Thus, the second research question in the current study examined children's beliefs about the three most common social class groups: rich, poor, and middle class.

Associations Between Children's Social Group Identification and Their Beliefs About Social Groups

Finally, as our third research question, we examined whether children's understanding and beliefs about social class differ by their own social class group membership, as some empirical evidence would suggest (e.g., Emler & Dickinson, 1985; Leahy, 1983). For example, Weinger (2000), in a qualitative study with middle-class and poor children, found that children described middle-class families in an idealized, positive way (e.g., good manners, happy, responsible), regardless of their own social class background. When asked to describe imagined poor families, however, the descriptions of study children whose families were poor conveyed better understanding of the financial hardships and social rejections experienced by poor individuals than did middle-class children's descriptions. Study children whose families were poor also exhibited greater empathy for the imagined poor family and were more likely to spontaneously express a desire to help the family as compared to their middle-class peers. Interestingly, children's motivations for wanting to be friends with a child from a background similar to their own also differed by social class background: Whereas middle-class children's preferences were based on perceived similarity (e.g., "Cuz they're like me"; Weinger, 2000, p. 143), the responses of children who were poor were more varied (e.g., "This kid [poor] would probably treat me nice and this kid [middle-class] would probably treat me badly because she's rich and she wants a lot of friends who are rich"; Weinger, 2000, p. 143). We are not aware, however, of any studies that have evaluated the extent to which children's notions about their own social class group informs their reasoning and beliefs about others on the basis of social class group membership.

Study Aims and Hypotheses

Informed by DIT (Bigler & Liben, 2006), this study explored 10- to 12-year-old children's (a) identification of and reasoning about their SSS; (b) beliefs about the poor, middle class, and rich; and (c) the associations between the two. Regarding the first research question, based on Goodman et al. (2000), we hypothesized that upper elementary school-aged children's SSS ratings would be correlated with both objective indicators of social class (i.e., parent education level, family income) and their parents' SSS ratings. In addition, as suggested by DIT, we expected that children's SSS would be based more on concrete and tangible features of their proximate environments, such as housing (e.g., neighborhood quality), material possessions (e.g., brand-name clothing, electronics), and life experiences (e.g., vacations), rather than references to abstract notions such as personal traits and evaluative judgments. Finally, considering that adults express a strong tendency to identify as middle class (Diemer & Ali, 2009; Evans & Kelley, 2004; Ladd & Bowman, 1998) and Goodman et al.'s observation that a majority of both upper-middle-class and working-class youth identified as middle class, we predicted that children would demonstrate a similar middle-class bias.

Regarding the second research question, based on DIT (Bigler & Liben, 2006), we expected that children would have distinct impressions of what it means to be rich and poor and would display evidence of essentialist beliefs about both groups. Consistent with prior studies (Sigelman, 2012; Woods et al., 2004), we expected that the children in this study would express the least favorable beliefs about the poor as compared to the rich and middle class. Furthermore, given the aforementioned tendency among adults and adolescents to identify as middle class and Weinger's (2000) observation of a privileging of middle-class status among children and early adolescents, we expected children's beliefs about the middle class to be more positive than their beliefs about the poor, and perhaps even the rich.

Finally, regarding our third research question, given that virtually no prior studies have examined links between children's subjective ratings of social status and their beliefs about members of other social class groups, we approached this as a more exploratory aim of the current study. It is possible, however, to offer two different (and competing) hypotheses. First, work within the DIT framework shows that children often develop biases favoring the in-group. As such, it is possible that lower SSS children will hold more favorable beliefs toward the poor than other children would (and hold more favorable attitudes toward the poor than other social class groups). It is just as likely, however, that lower SSS children will distance themselves from the poor, rating them as more negative than other children. This may reflect, in part, knowledge of the lower social status of being poor and children's desire to maintain their own positive sense of self distinct from the unfavorable group (see Ellemers & van Laar, 2010) or system-justifying beliefs that maintain the status quo (see Flanagan et al., 2014; Jost, Pelham, Sheldon, & Ni Sullivan, 2003).

A distinguishing feature of the current study is its mixed-methods research design. Mixed-methods research is "the class of research where the researcher mixes or combines quantitative and qualitative research techniques, methods, approaches, concepts or language into a single study" (Johnson & Onwuegbuzie, 2004, p. 17). Mixed-methods research is valued because it affords greater methodological flexibility and inclusivity. A key advantage of such an approach is the use of multiple data-gathering strategies and approaches, concurrently or sequentially, that permit examination of convergence and corroboration of findings across data streams more so than is possible when using a monomethod research approach. Doing so can help to generate insights that might be missed via a monomethod approach and can help to better inform theory and practice (Creswell & Plano Clark, 2007).

In the current study, quantitative data yielded ratings of children's SSS and beliefs about social class groups, and allowed us to assess relations between SSS and attitudes about others on the basis of social class group membership. Qualitative data sources allowed us to more thoroughly probe children's beliefs about social class, including the sources of evidence (e.g., personal experiences) and social markers (e.g., material possessions) that inform their perceptions and beliefs. This process yielded a rich description of children's constructions of their SSS and beliefs about social class groups.

Method

Research Site and Participants

Data for the current study were collected at a university-affiliated, community-based, laboratory school in Los Angeles County, California. Approximately 450 students from diverse SES backgrounds attend the school. Although the school is tuition based, roughly 25% of students receive tuition support. Annual household incomes at the school range from < $10,000 to more than $1 million; median income was in the $150,000–$199,999 range at the time of data collection. Though this is a higher median income than that of Los Angeles County ($56,241 for the 5-year period from 2008 to 2012; U.S. Census Bureau, 2012), it is on par with the neighborhoods most proximate to the school (range = $74,830–$141,527; http://zipatlas.com). The numbers should also be considered within the context of Los Angeles County, 1 of the 10 most expensive counties in the nation (http://www.kiplinger.com). The student body at the school was also racially and ethnically diverse: 32% of students were Euro-American, 23% Latin American, 10% African American, 10% Asian American, 4% American Indian, 18% Multiethnic, and 3% Other.

A total of 117 upper elementary students (fifth and sixth grades) participated in the current study (49% male; see Table 1 for sample chararcteristics and descriptive statistics). Parent data ($N = 112$) were collected via survey packets sent home with child consent forms. Consistent with the schoolwide demographics, families of participating students were primarily middle to upper middle class and, based on parent report, 33% were Euro-American, 21% Latin American, 6% African American, 11% Asian American, and 30% Multiethnic.

Table 1
Descriptives for Study Sample

	%	Range	*M*	*SD*
Child characteristics				
Gender				
Female	51.30			
Male	48.70			
Race and ethnicity				
White	32.50			
Latin American	21.40			
African American	5.10			
Asian American	11.10			
Multiethnic	29.90			
Subjective social status				
Parent report		1–10	6.07	2.33
Child report		3–10	5.86	1.44
Socioeconomic status				
Household income			8.07	3.09
(1) < 10 K	0.90			
(2) > 10–15 K	2.80			
(3) > 15–25 K	0.90			
(4) > 25–50 K	12.80			
(5) > 50–75 K	9.20			
(6) > 75–100 K	6.40			
(7) > 100–150 K	11.90			
(8) > 150–200 K	6.40			
(9) > 200–250 K	7.30			
(10) > 250–500 K	11.90			
(11) > 500–750 K	10.10			
(12) > 750 K	19.30			
Maternal education level			4.84	1.51
(1) Did not graduate HS	5.10			
(2) HS graduate	9.10			
(3) Vocational school	5.10			
(4) Community college	2.00			
(5) College	35.40			
(6) Postgraduate degree	43.40			
Paternal education level			4.84	1.67
(1) Did not graduate HS	7.60			
(2) HS graduate	10.90			
(3) Vocational school	1.10			
(4) Community college	3.30			
(5) College graduate	25.00			
(6) Postgraduate degree	52.20			
Beliefs about people who are poor, middle class, and rich				
Attitude ratings: positive				
Ratings of the poor		1.50–5.00	3.18	0.66
Ratings of the middle class		2.83–5.00	3.65	0.49
Ratings of the rich		1.25–5.00	3.62	0.55
Attitude ratings: negative				
Ratings of the poor		1.00–5.00	2.58	0.60
Ratings of the middle class		1.00–3.60	2.30	0.49
Ratings of the rich		1.00–3.75	2.41	0.53

Note. *N* = 117 students, 112 parents. HS = high school.

Procedures and Measures

Overview

Children were individually interviewed in a private room at the school. Children responded orally to all open-ended questions and completed survey items on their own, with the interviewer present. Interviews were audio-recorded and transcribed by members of the research team. Children were asked group identification questions (i.e., SSS questions) first and attitudinal questions in a latter portion of the interview protocol.

Subjective Social Status

Children's and parents' SSS were assessed using a modified version of a measure used by Goodman et al. (2001). Modifications were informed by previous research and extensive piloting. Given that children of this age typically define wealth and poverty in monetary terms, we focused on this dimension of SES. Children were shown a picture of a 10-rung ladder while the interviewer read the following:

> Imagine that this ladder pictures how American society is set up. At the top are the people who have the most money and at the bottom are the people who have the least money. Now, think about your family. Where do you think they would be on this ladder? Mark an X on the step where your family would be on this ladder.

The location of each child's X was coded such that 1 represented the bottom rung of the ladder and 10 represented the top rung. We also assessed what informs children's SSS by asking: "Why did you place your family here on the ladder?" Parents were asked the identical SSS question in survey form.

Socioeconomic Status

Parents used a scale ranging from 1 (< *$10,000*) to 12 (*more than $750,000*) to indicate their family's household income based on all sources. Parents also indicated their highest level of education, on a scale that ranged from 1 (*did not graduate high school*) to 6 (*postgraduate training completed*). Average reported income was 8.07 (*SD* = 3.09), which corresponds to between $150,000 and $200,000. Parent education ranged from less than high school to professional degrees, but average parent education was high: Over 75% of parents had a college degree or higher.

Due to the often-observed confound between SES and racial and ethnic background, we conducted preliminary analyses to examine racial and ethnic differences in participant families' SES. An analysis of variance indicated that household incomes differed by child's race and ethnicity, $F(4, 104) = 8.05$, $p < .001$, $\eta^2 = .24$, but it is important to note that there was substantial variability in family income within each group, as evidenced by the standard deviations (see Table 1). Bonferroni analyses (used throughout) indicated that Latin American children's family income was lower ($M = 5.70$, $SD = 3.02$) than their Euro-American ($M = 9.83$, $SD = 2.61$, $p < .001$) and Multiethnic ($M = 8.15$, $SD = 2.55$, $p = .013$) peers, who did not differ from each other. African American ($M = 7.20$, $SD = 3.27$) and Asian American children's ($M = 7.67$, $SD = 3.00$) family incomes fell between these ranges and did not differ from those of other children. Similar patterns of difference by child's ethnicity were observed for maternal education, $F(4, 94) = 3.65$, $p = .008$, $\eta^2 = .13$, and paternal education, $F(4, 87) = 5.13$, $p = .001$, $\eta^2 = .19$, with differences particularly pronounced between parents of Latin American children and Euro-American children (full results available upon request).

Beliefs About People Who Are Poor, Middle Class, and Rich

During the interview, we assessed children's beliefs qualitatively by asking them to respond to the following:

> I'm going to read some words and phrases to you and I want you to tell me whatever comes to your mind when I read those words. What do you think of when you hear the word or phrase . . . [poor, middle class, rich]?

We also assessed beliefs quantitatively using a trait-stereotyping measure similar to the multiresponse racial attitude (Doyle & Aboud, 1995) and the Preschool Racial Attitude Measure II (Williams, Best, Boswell, Mattson, & Graves, 1975). Comparable adaptations of these measures have been used in other studies with elementary age children (see Bigler, Jones, & Lobliner, 1997; Pfeifer et al., 2007). Using a 1 (*none*) to 5 (*almost all*) scale, children indicated how many people from three social class groups (poor, middle class, rich) exhibited each of the following attributes: smart, hardworking, clean, good, honest, polite, dumb, lazy, dirty, bad, liar, rude. Consistent with prior studies, and based on results from exploratory factor analyses using principal axis factoring with promax rotations (conducted for each social class group separately), two scores were created for each of the three social class groups by averaging responses for each of the six positive attributes (smart, hardworking, clean, good, honest, polite) and six negative attributes (dumb, lazy, dirty, bad, liar, rude) separately. Results of the exploratory factor analyses and individual item means and standard deviations are reported in Table 2. Factor analyses were run using two criteria: first, specifying an eigenvalue ≥ 1, and second, based on our a priori expectations, specifying a two-factor structure. Both specifications yielded a consistent two-factor structure for beliefs about middle-class attributes. For attributes about the poor and rich, the minimum eigenvalue specification yielded a three-factor solution in which one or two items loaded on a standalone factor whereas the two-factor specification yielded a solution consistent with attributes about the middle class. To ensure a comparable factor structure across social class groups, and consistent with prior studies, we elected to create scale scores based on the two-factor solution. Cronbach's alphas for ratings of the poor, middle class, and rich, respectively, were: .82, .81, and .78 for positive attributes, and .76, .82, and .79 for negative attributes.

Analytic Strategy

A concurrent triangulation mixed-methods design was used in which complementary quantitative and qualitative data were collected to address the study's aim (Creswell & Plano Clark, 2007). The distinct forms of data were weighted equally and merged during the data analysis and interpretation phase. Such designs are useful when the goal is to inform a single overall interpretation of the study findings (Creswell & Plano Clark, 2007). For the purposes of the current study, quantitative data sources included SSS ratings, family SES indicators (i.e., income, parent education level), and survey data assessing children's stereotypic beliefs about the poor, middle class, and rich. Qualitative data sources included children's justifications for their SSS ratings and beliefs about the poor, middle class, and rich based on a spontaneous association task. As described in the Results section below, we used several different strategies to analyze these data and answer our research questions.

Results

Research Question 1: How Do Children Identify Their Own Social Class Group? What Informs Their Identification?

To examine if children identify with a social class group and what informs their identification, we

Table 2

Factor Loadings for Children's Beliefs About the Poor, Middle Class, and Rich

		Eigenvalues ≥ 1			Two-factor solution	
Items	*M* (*SD*)	Factor 1 positive attributes	Factor 2 negative attributes	Factor 3 miscellaneous	Factor 1 positive attributes	Factor 2 negative attributes
How many poor people are (*n* = 96):						
Smart	3.06 (0.72)	**0.55**	−0.30	−0.23	**0.55**	−0.32
Good	3.48 (0.83)	**0.71**	−0.47	−0.34	**0.72**	−0.51
Polite	3.24 (0.91)	**0.73**	−0.40	−0.12	**0.72**	−0.39
Honest	3.02 (0.70)	**0.66**	−0.46	−0.25	**0.67**	−0.47
Clean	2.63 (0.99)	**0.72**	−0.25	−0.32	**0.70**	−0.33
Hardworking	3.47 (1.14)	**0.64**	−0.45	−0.19	**0.65**	−0.43
Liars	2.48 (0.84)	−0.38	**0.73**	0.33	−0.38	**0.70**
Dirty	2.99 (1.07)	−0.59	**0.41**	0.33	−0.59	**0.46**
Lazy	2.54 (0.95)	−0.26	**0.46**	0.08	−0.26	**0.39**
Dumb	2.56 (0.74)	−0.33	0.39	**0.91**	−0.33	**0.50**
Bad	2.41 (0.75)	−0.50	**0.72**	0.46	−0.51	**0.75**
Rude	2.38 (0.73)	−0.31	**0.54**	0.44	−0.31	**0.60**
	Eigenvalue	4.60	1.44	1.07	4.60	1.44
	% variance	38.35	12.03	8.90	38.35	12.03
How many middle-class people are (*n* = 95):						
Smart	3.74 (0.60)	**0.59**	−0.52	—	**0.59**	−0.52
Good	3.66 (0.63)	**0.78**	−0.40	—	**0.78**	−0.40
Polite	3.58 (0.66)	**0.82**	−0.26	—	**0.82**	−0.26
Honest	3.39 (0.74)	**0.58**	−0.33	—	**0.58**	−0.33
Clean	3.69 (0.74)	**0.56**	−0.27	—	**0.56**	−0.27
Hardworking	3.77 (0.68)	**0.64**	−0.39	—	**0.64**	−0.39
Liars	2.46 (0.69)	−0.28	**0.51**	—	−0.28	**0.51**
Dirty	2.09 (0.63)	−0.39	**0.55**	—	−0.39	**0.55**
Lazy	2.47 (0.73)	−0.24	**0.74**	—	−0.24	**0.74**
Dumb	2.17 (0.66)	−0.44	**0.76**	—	−0.44	**0.76**
Bad	2.26 (0.59)	−0.35	**0.64**	—	−0.35	**0.64**
Rude	2.33 (0.63)	−0.34	**0.80**	—	−0.34	**0.80**
	Eigenvalue	1.81	4.72	—	1.81	4.72
	% variance	15.06	39.36	—	15.06	39.36
How many rich people are (*n* = 98):						
Smart	3.70 (0.77)	**0.60**	−0.32	−0.13	**0.60**	−0.33
Good	3.53 (0.74)	**0.74**	−0.32	−0.18	**0.73**	−0.36
Polite	3.65 (0.72)	**0.70**	−0.32	−0.38	**0.72**	−0.40
Honest	3.27 (0.71)	**0.62**	−0.34	−0.18	**0.62**	−0.36
Clean	4.03 (0.91)	0.48	−0.14	**−0.62**	**0.47**	−0.27
Hardworking	3.47 (0.89)	**0.57**	−0.39	−0.09	**0.56**	−0.39
Liars	2.62 (0.75)	−0.30	**0.60**	0.27	−0.33	**0.61**
Dirty	1.84 (0.74)	−0.11	0.40	**0.92**	−0.22	**0.41**
Lazy	2.80 (0.92)	−0.29	**0.79**	0.22	−0.32	**0.76**
Dumb	2.26 (0.69)	−0.55	**0.53**	0.31	−0.57	**0.57**
Bad	2.34 (0.68)	−0.52	**0.75**	0.37	−0.55	**0.78**
Rude	2.63 (0.77)	−0.43	**0.64**	0.20	−0.45	**0.64**
	Eigenvalue	4.49	1.54	1.30	4.49	1.54
	% variance	37.42	12.84	10.83	37.42	12.84

Note. Principal axis factoring with a promax rotation. Factor structure denoted by bolded factor loadings.

first computed descriptive statistics for our sample's SSS. Then, we used bivariate correlations to examine how children's SSS ratings were associated with indicators of SES (e.g., family income) and parents' SSS ratings. Finally, to assess what informed children's SSS ratings, we coded their open-ended responses in an inductive manner and categorized their responses according to major themes.

Descriptives

As shown in Table 1, children used almost the full range of the ladder (range = 3–10) when assessing their SSS. The average SSS was 5.86 (*SD* = 1.44). For group difference analyses (reported below), we grouped children into three categories: 17 children had lower SSS ratings (ladder ratings of 3 and 4), 73 children rated themselves in the middle (ladder ratings between 5 and 7), and 11 children had higher SSS ratings (between 8 and 10). We intentionally avoid labeling these groups as "poor," "middle class," or "rich" because we cannot assume that any particular SSS corresponds with identifying with one of these terms.

Associations With SES and Parents' SSS

Children's SSS ratings were positively correlated with all three SES indicators: household income ($r = .64$, $p < .001$), mother's education level ($r = .23$, $p = .034$), and father's education level ($r = .29$, $p = .009$). These correlations remained essentially the same after controlling for the child's ethnicity (dummy coded; 1 = Latin American). Thus, children's SSS ratings reflect their families' socioeconomic position.

Next, we examined the correspondence between children's and parents' SSS ratings. Parents' SSS ratings ranged from 1 to 10, with an average of 6.07 (*SD* = 2.33). As with children, parents' SSS was positively correlated with household income ($r = .80$, $p < .001$) and maternal ($r = .42$, $p < .001$) and paternal ($r = .47$, $p < .001$) education level. Children's SSS rankings were positively associated with their parents' SSS ($r = .56$, $p < .001$). This association held even after controlling for the child's ethnicity (partial $rs = .55$, $p < .001$) and parents' education level (partial $r = .44$ and .55, $p < .001$, paternal and maternal education levels, respectively), but not household income (partial $r = .09$, *ns*). Thus, while SSS is not entirely synonymous with objective indicators of SES, the overlap between parents' and children's SSS ratings is more strongly informed by income than parental education level. This makes sense given that the stem of the SSS prompt included reference to the amount of money that one's family has.

Social Markers That Inform Children's SSS

To examine the social markers that children used to inform their SSS, we used an iterative process to code children's open-ended responses to the probe: "Why did you place your family here on the ladder?" During the code development phase, research team members read random subsets of children's responses (i.e., random pulls of 25% of the responses) and developed an exhaustive set of categories, which were then aggregated into five summary categories. Once no additional codes were identified and the team was able to reach consensus on coding of the random subsets, two team members independently coded all of the responses into these five categories and resolved disagreements by consensus. Disagreements were not tracked at this final stage of coding. Coding was done at the statement level; that is, each statement was coded using all applicable codes.

The five categories (and the percentage of responses that were coded as these categories) were: (a) *purchasing power* (47%): references to money as a resource and how money is used; (b) *social references* (29%): comparisons to other points on the SSS ladder or SES spectrum, comparisons with other individuals, or reflections upon interpersonal and intrapersonal experiences; (c) *source of income, wealth, and assets* (13%): references to where money comes from (e.g., parents' occupations and family assets); (d) *psychological affect* (9%): references to the emotions, judgments, or values associated with having (or not having) access to material resources, or money; and (e) *structural factors* (2%): references to societal, institutional, or family sociodemographic characteristics. On average, children provided two uniquely coded responses (*SD* = 1.77, range = 1–7).

References to purchasing power were by far the most common; nearly half of all responses included a reference to either housing ("we live in an okay house . . . it's big and spacious"; "we live in a really tiny apartment"), material possessions ("we have enough money to, you know, pay for our needs"; "my dad, he has a more expensive car"), or spending habits, affordances, and restrictions ("we can afford extras"). Responses displayed an awareness of the multiple ways in which money manifested itself in children's daily experiences. For example, a child who rated himself a 7 on the ladder explained,

> Because I think we have a pretty nice house and we get some of the things that we want and it's pretty luxurious at my house so I think we have a pretty good amount of money but we don't have like the most money. . . . The schools I go to . . . some of the camps . . . . And the trips we go on a lot, we've gone to like Tokyo, Africa, Fiji, Tahiti.

Spending habits featured prominently in children's estimations of their SSS, as evident in the following quote,

> . . . we're middle class . . . my parents try to use money wisely, so it makes me feel around 6 . . . my mom, she tries to get lots of sales. She tries to find good gas at good prices. And she tries to carpool when possible.

For some, SSS was informed by restricted access to resources or activities. For example:

> I have to share a room with my sister, and my mom and dad have to share a room with my other sister. . . . We are a family of five, and there has been times when my dad has lost his job and we haven't been having food. (SSS = 3)

Children's responses also indicated sensitivity to fluctuations in purchasing power due to social and economic conditions. One child remarked, "I picked [4] because right now, I mean, we're renting so that's normal but the renting price is going up and the gasoline prices are going way up so we're a little low on money."

The next most frequently occurring response category included social references either to explicit others or to more general perceptions of social status. In some instances, comparisons were fairly general such as, "We're not rich, and we're not poor. We're right in the middle" (SSS = 4). In other instances, the comparisons were quite specific and included reference to public figures ("Bill Gates wealthy"), marked social groups such as "the homeless" or "millionaires," or someone known to the child such as friends or peers at school. As shown in the following example, children's responses sometimes demonstrated a sophisticated and nuanced awareness of status,

> I probably wouldn't have [known] if . . . I don't live where I do, but . . . most of my friends who live in my neighborhood, they're all very wealthy and . . . their houses are very big and they have pets and all this stuff, and then I live in a condominium, and, even though it's the same neighborhood, it's like two different parts . . . some of my friends, I'll go to their house and their house will also be very fancy, . . . so it's mostly compared to other people. But if I were to just not have been to any of those places, I would say we have a good amount of money. (SSS = 3)

Another child commented,

> Well, I was thinking that at the top of the ladder, that's someone like J.K. Rowling, they'd just be really wealthy and the next one is like someone who is pretty wealthy and this one is [pointing to the next rung on the ladder], they probably live in a really big house and I think I might be here [points down] because I think my family has enough money to be comfortable and we're happy and like we don't have a hard time, but it's not like we have a lot of money and it's not like we live in a huge house—we live in a little house in [neighborhood] but it's just nice and a nice size. (SSS = 7)

Interestingly, some children indicated that they were at times the target of social comparisons, and the attention was not always welcomed. For example, one child (SSS = 8) said,

> Well, some of my friends, tons of my friends actually, they're like "oh wow, you get to do that, my family can't afford that." And it doesn't make me feel good. It sort of makes me feel like, well, I sort of feel bad for people and stuff.

Another child (SSS = 7) discussed how she tried to downplay her family's affluence in the presence of friends whom she knew to have a lot less than she did,

> Well maybe like [put my expensive toys] in a drawer or something or say oh let's do something else or something like that. . . . I don't want them to feel uncomfortable because they realize I'm really wealthy and they're not as wealthy as me or something.

A smaller percentage of children's responses focused on the *source* of money and wealth. Children made reference to their parents' occupations (e.g., brain surgeon, stylist, writer, "my dad owns a kayak business"), job status ("my mom has to work two jobs"; "there have been times when my dad lost his job"), and salary (e.g., "because my mom has a really good job . . . she makes a lot of money"). In general, many children demonstrated specific knowledge of parents' occupations. For example,

> I think we're on the moreish money side 'cuz my mom's a rheumatologist and lots of people need rheumatologists now, and my dad's a pediatrician and so they get like a lot of money from doing the things that they do. . . . (SSS = 6)

For others, SSS was informed by more of an implicit awareness of one's family finances ("we don't have a lot of money . . . We just don't. I just know"), awareness of current wealth ("my mom always owned a house . . . and then my dad owns a house in Kauai, so I mean, if they were really in trouble, they could sell a house"), or awareness of inherited wealth (". . . my dad comes from a family with like a lot of money and stuff").

For some children, discussions of the source of social and economic mobility were intertwined with stories about their family heritage, as highlighted in the following examples.

> [My family] first got here being immigrants so when they got their jobs they weren't like really at a big job. My dad did get a big job [selling produce to restaurants across the state] because he's been here since he was 15 so he has been working and like getting higher and higher but my mom, she got here at twenty-six I think so she at first didn't work . . . and then she started working cleaning houses and then she became [a] babysitter so I don't think they're very, very low in money, they're just in the middle. (SSS = 4)

Another child stated,

> My grandpa was born over in Mexico and he came here when he was about five. . . . and they were working on railroads and they were barely having enough to eat and he got, and then his brother came up with my dad's business which has lasted the family over I think three generations now. . . . I think [my grandfather is] one of the richest Latin Americans . . . as a kid being at barely surviving, now he lives in [an affluent neighborhood] in a mansion with a swimming pool. . . . (SSS = 7)

The two categories children cited least often were psychological affect (i.e., emotions) and structural factors (such as family or societal characteristics). When emotions were mentioned, they were often positive ("I'm really happy for all what we have"), but were sometimes negative ("I really don't feel comfortable answering this because my family had a really hard time since the writer's strike, my mom couldn't get a job so things have been really hard for her"). The few structural factors that were mentioned were generally about the economy as a whole ("You know how the bank is like losing money?").

To summarize, we found that in justifying their SSS, the most common explanations were references to possessions and lifestyle, including where children lived and their experiences (e.g., vacations, summer camps). In addition, we see that by the upper elementary grades, children are perceptive of themselves in relation to others; SSS ratings were informed by comparisons to others. Finally, the children in this study displayed awareness about key markers of social status such as jobs, income, and wealth, but rarely mentioned structural factors or affective characteristics.

As a final step, we examined children's explanations across SSS groups. Across all three SSS groups, the most frequently mentioned category was purchasing power (47% of 40 responses for lower SSS groups, 47% of 217 responses for middle SSS groups, and 39% of 33 responses for higher SSS groups) followed by social comparative statements (22%, 31%, 21% for lower, middle, and higher SSS groups, respectively). Higher SSS children mentioned income (24%) more so than did children in the lower (15%) and middle (10%) SSS groups. Psychological affect and structural factors were rarely mentioned by children of any SSS (psychological affect: 5%, 9%, 13% and structural factors: 8%, 1%, 0%, for lower, middle, and higher SSS groups, respectively).

Research Question 2: What Are Children's Beliefs and Stereotypes About the Poor, Middle Class, and Rich?

Children's Spontaneous Associations With Terms Poor, Middle Class, *and* Rich

Children's spontaneous associations with the terms *poor, middle class, and rich* were coded using an iterative process. Research team members read random subsets of children's responses and developed an exhaustive set of eight categories, which were then aggregated into three broad and mutually exclusive categories: *descriptive phrases and words*, *perceptually salient characteristics*, and *inferred characteristics*. Two team members then independently coded half of the responses. To establish interrater reliability, 25% of all of the responses were double-coded. Cohen's Kappas were .90, .78, and

Table 3
Children's Spontaneous Associations With Terms Poor, Middle Class, *and* Rich

		Percent of children who describe:		
Themes	Example	Poor	Middle class	Rich
Descriptive phrases				
Definitional	*Don't have a lot of money; not rich, not poor, in the middle*	42_a	52_a	40_a
Reference	*Homeless person; third world country; me; my family; most of the people I know at this school; suburbs; Bill Gates*	44_a	43_a	25_b
Perceptually salient characteristics				
Material possessions	*Clothes are messy; small home; regular or comfortable home; they have enough stuff; mansions; expensive cars; fancy clothes*	21_a	16_a	46_c
Occupation	*Owns like a big business; nice jobs*	5_a	0_b	11_a
Inferred characteristics and associations				
Adjectives	*Really nice; very successful; very generous; not nice; snobby*	19_a	11_a	18_a
Quality of life	*Don't have enough money to pay bills; enough money to be comfortable; pretty much not worry about anything financial*	22_a	30_b	22_a
Affective response	*Feel bad for them; they should not be out on the streets; I'd like to help them; it doesn't make them not a good person*	27_a	5_b	7_b
Within-group heterogeneity	*I think of people who buy big houses and boats and expensive cars . . . but I think there's also people who do have the same amount of money but don't waste it. I think there's two categories of poor people–the people who are doing fine and are satisfied even though they don't have that much money and people who are really just not doing well*	3_a	1_a	9_b

Note. Repeated measures multinomial logistic regressions conducted to test for differences in children's conceptions of poor, middle class, and rich. Analyses covary influence of child's gender (1 = boy) and ethnicity (1 = Latin American), and household income. Differing subscripts denote differences between groups at $p < .05$; same subscripts denote no difference. Detailed results available upon request.

.77 for children's responses to poor, middle class, and rich, respectively.

The first broad category, descriptive phrases and words, includes the subcategories definitional (such as describing how much money a person has or giving a synonym for wealth and poverty) and social references (describing individuals or referencing media). The second broad category, perceptually salient characteristics, includes mention of material possessions (e.g., housing, clothing, cars, entertainment items, and toys) and *occupations*. Finally, the third broad category, inferred characteristics, includes the use of adjectives, describing quality of life (e.g., mentioning lifestyle choices and opportunities), invoking affective terms such as sympathy, empathy, or emotional reactions to having money or to inequality in the United States, and within-group heterogeneity (i.e., explicit mention of heterogeneity within a particular social class). Table 3 lists examples of responses for each category and shows the sample distribution across the three social class groups.

To determine whether children's associations differed systematically by the target group (poor, middle class, rich), we ran separate repeated measures multinomial logistic regressions for each of the eight micro categories shown in Table 3. All analyses controlled for children's gender (1 = boy), ethnicity (1 = Latin American), and household income. Each regression was run twice: once with rich as the reference group, and again with poor as the reference group. Results revealed that children's conceptions of social class were different for the poor, middle class, and rich. Although children were equally likely to give definitional and adjective responses to the terms poor, middle class, and rich, children's use of the other categories (i.e., social reference, material possessions, occupation, quality of life, affective responses, and within-group heterogeneity) varied as a function of the social class they were discussing. Specifically, poor and middle class were more likely than rich to invoke references to individuals and media. It is interesting to note that self-references (e.g., "my family, friends," "I see people like me"), which were included as social references for analysis purposes, were strikingly different by target group: Whereas very few children referenced themselves or their family when responding to the terms *rich* and *poor* (*ns* = 3 and 2, respectively), many more children did so in response to the term *middle class* ($n = 26$). Furthermore, of the two children who included a self-reference when responding to the term *poor*, both did so in a socially distancing manner (i.e., "not me"; ". . . I'm middle class, so like, I'm not poor but I'm not rich").

Affective or empathetic responses were more common when children were discussing their conceptions of poor. Lifestyle and quality of life expressions were more frequent in children's descriptions of the middle class. Finally, children were more likely to mention material possessions and within-group heterogeneity when describing the rich as compared to either the middle class or poor.

Children's Beliefs About the Poor, Middle Class, and Rich

We conducted a repeated measures analysis of covariance (ANCOVA) to examine differences regarding the extent to which children endorse positive and negative attributes about people who are poor, middle class, and rich. This model included two within-subject factors: target group (poor, middle class, and rich) and valence of attribute (positive and negative) and three covariates: children's gender (1 = boy), ethnicity (1 = Latin American), and household income.

There was a significant interaction between target group and attribute valence, $F(2, 178) = 5.84$, $p = .003$, partial $\eta^2 = .06$. Follow-up ANCOVA pairwise comparisons indicated that children endorsed fewer positive attributes about the poor than they did about either the middle class, $F(1, 91) = 11.24$, $p = .001$, partial $\eta^2 = .11$, or the rich, $F(1, 92) = 3.70$, $p = .058$, partial $\eta^2 = .04$ (see Figure 1). Children's endorsement of positive attributes about the middle class and the rich did not differ from one another, $F(1, 92) = 2.64$, $p = .108$, partial $\eta^2 = .03$. Differences were also observed in children's endorsement of negative attributes. Specifically, children endorsed more negative attributes about the poor than they did about the middle class, $F(1, 89) = 5.29$, $p = .024$, partial $\eta^2 = .06$. Children's endorsement of negative attributes about the rich did not differ from those of either the poor, $F(1, 90) = 1.86$, $p = .176$, partial $\eta^2 = .02$, or the middle class, $F(1, 90) = 0.67$, $p = .414$, partial $\eta^2 = .01$.

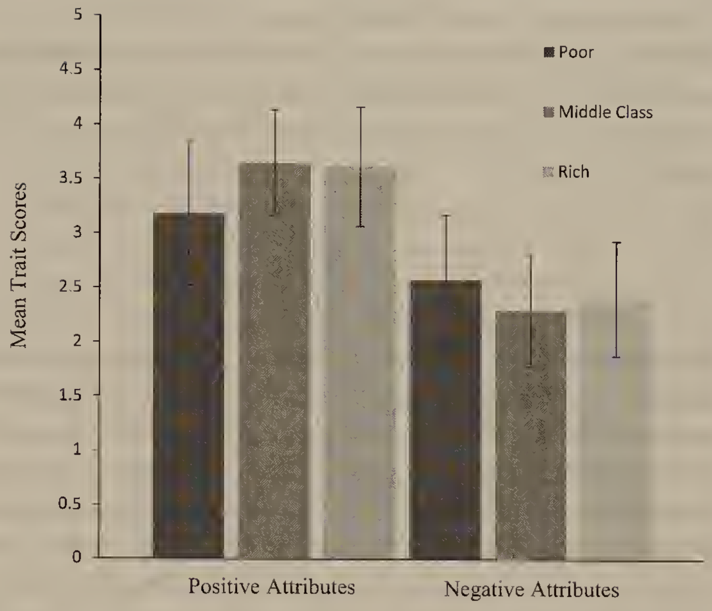


Figure 1. Children's beliefs about the poor, middle class, and rich ($N = 117$; $ns = 98$ to 101).
Note. Shown with standard deviation bars.

Research Question 3: Do Children's Beliefs About the Poor, Middle Class, and Rich Vary by SSS?

Building on our earlier analyses, we conducted a mixed-factorial ANCOVA to answer our third research question. This model included (a) two within-subject factors: target group (poor, middle class, and rich) and valence of attribute (positive and negative) based on the trait stereotype scores; (b) one between-subject factor: children's SSS grouping (low, middle, and high); and (c) three covariates: gender (1 = boy), ethnicity (1 = Latin American), and household income. There was a significant three-way interaction between target group, attribute valence, and SSS group, $F(4, 172) = 2.96$, $p = .021$, partial $\eta^2 = .06$.

To better understand the source of the three-way interaction, we ran separate mixed-factorial ANCOVAs for children's positive and negative attributes. For children's endorsement of positive attributes, the interaction between target group and SSS was not significant, $F(4, 176) = 1.85$, $p = .121$, $\eta^2 = .04$, indicating that children's endorsement of positive attributes about the poor, middle class, and rich did not differ by the child's own SSS. For negative attributes, however, the interaction between target group and SSS was significant, $F(4, 172) = 2.44$, $p = .049$, $\eta^2 = .05$. Follow-up ANCOVA pairwise comparisons indicated that low-SSS children endorsed significantly more negative attributes about the poor ($M = 2.94$, $SD = 0.76$) than did their middle-SSS peers ($M = 2.50$, $SD = 0.56$), $F(1, 80) = 4.59$, $p = .035$, partial $\eta^2 = .06$. Although the low-SSS children differed even more from their high-SSS peers ($M = 2.48$, $SD = 0.32$) than they did from their middle-SSS peers, the difference in their negative attributes for the poor was not statistically significant, $F(1, 26) = 1.18$, $p = .289$, partial $\eta^2 = .05$.

This nonsignificant difference was likely due to low power. We observed no difference between middle- and high-SSS children in their negative attitude trait scores toward the poor, $F(1, 75) = 0.03$, $p = .861$, partial $\eta^2 = .00$.

Discussion

Guided by DIT (Bigler & Liben, 2006) and prior studies examining youths' and adults' conceptions and beliefs about social class, the current mixed-methods study had three primary research questions: (a) How do 10- to 12-year-old children identify and reason about their SSS? (b) What are their beliefs about the poor, middle class, and rich? (c) What is the association between their SSS and their beliefs about social class groups? Findings suggest that children, by late elementary school, understand their own social class in ways that parallel adolescents and adults, albeit with more of a focus on the concrete features of social class.

Specifically, for the first research question, the findings supported our hypothesis. First, children's SSS ratings were correlated with both objective indicators of social class (i.e., parent education level, family income) and parents' SSS. This is consistent with previous research (e.g., Adler et al., 2000; Goodman et al., 2000; Goodman et al., 2001) but extends this finding to preadolescent children. Second, also consistent with prior studies with adults, nearly three fourths of the children in the current study rated their SSS as somewhere in the middle (i.e., ladder ratings between 5 and 7) and many children, regardless of their SSS rating, invoked a reference to themselves when giving a spontaneous response to the term *middle class*. Like adults, children seem to consider themselves as part of the normative group. It may be that this tendency to identify as middle class is due to children envisioning extreme examples of rich (e.g., Bill Gates) and poor (e.g., homeless individuals), and thus falling somewhere between those extremes implies middle class. It may also be the result of children perceiving that middle class is "normal" or "regular" (Weinger, 2000) and children are simply motivated to feel typical. Future research should further explore whether cognitive or motivational biases drive this tendency in children.

Furthermore, as expected, qualitative data revealed that children's SSS was informed by both features of and experiences in their proximal environments. Consistent with DIT (Bigler & Liben, 2006), children were particularly focused on the concrete and tangible components of their social class, such as what they are (or are not) able to buy with their family's income, what their house is like, and how their possessions and lifestyle differ from their friends. Their conceptions were also informed by what they see in the media (e.g., when they reference celebrities like J. K. Rowling). These findings support DIT's assertion that children make specific group–attribute links based on their exposure to salient environmental messages. Overall, the finding that children focus on the concrete manifestations of social class parallels children's understanding of other social groups, such as race and ethnicity (Quintana, 1998), in which children of this age focus on concrete racial differences (such as skin color) before they focus on more abstract understandings of race.

For our second research question, the results were largely consistent with our hypothesis. As predicted, quantitative data revealed that children did associate the least favorable attributes with the poor relative to the rich. They rated the poor as both higher in negative attributes and lower in positive attributes compared to the other social class groups. These findings are consistent with prior studies showing negative appraisals of the poor among adolescents (Flanagan & Tucker, 1999; Flanagan et al., 2014; Mistry, Brown, Chow, & Collins, 2012; Skafte, 1989). Ironically, although the poor were perceived the least favorably of all social class groups, children's descriptions of the poor often involved expressions of sympathy. This is consistent with prior studies with elementary school-aged children (Chafel & Neitzel, 2005).

Beyond the poor, children did not differentiate between the rich and the middle class in their quantitative ratings of either positive or negative attributes. This does not suggest, however, that children do not differentiate between these two groups. Rather, their spontaneous associations reveal that children often linked the rich with material possessions, in contrast to referencing themselves when describing the middle class. Children further clarified an understanding of the rich by differentiating between two types of rich. There are . . . "people who buy big houses and boats and expensive cars" and there are those who are more socially conscious, "who do have the same amount of money but don't waste it."

As mentioned previously, many children appeared to focus on the extreme ends of the economic spectrum in their spontaneous associations. For example, when discussing the poor, children mentioned the homeless, and more specifically, homeless adults (not families and children). Likewise, when thinking about the rich, children referenced high-profile celebrities such as J. K. Rowling. This is consistent with developmental intergroup research that shows that the most prototypical, or stereotypical, examples of social group members are the most salient to children (Brown & Bigler, 2005). Our findings are also consistent with prior research indicating elementary-age children's tendency to associate wealth with an abundance of material possessions and poverty with the total absence of material goods, while simultaneously showing some ambivalence toward the rich and greater sympathy toward the poor (Chafel, 1997; Chafel & Neitzel, 2005; Flanagan et al., 2014; Weinger, 2000).

For the third and final research question, linking children's attitudes about social class groups with their own SSS, one of our competing hypotheses was partially supported. Specifically, children who perceived themselves lower in SSS (i.e., 3 or 4 on the 10-point SSS ladder measure) rated the poor as having more negative attributes than the other more middle SSS children (i.e., 5–7). In interpreting this association, it is important to note two caveats. First, beyond overall small effect sizes, the average scores for all three social class groups (i.e., poor, middle class, rich) indicated that between "few" and "some" people displayed the traits in question (see Table 1). Second, none of the study participants rated themselves on the two lowest rungs of the ladder (although a few parents did) and none self-identified as poor. Thus, while it appears that some children were cognizant of their overall lower social status relative to others, it cannot be assumed that children in the lower SSS group identified as a member of the group that they were assessing (as was true for those who rated their SSS in the middle).

However, as predicted, given children's awareness of the pervasive negative societal stereotypes about the poor, children who rated themselves lower in terms of SSS may have been motivated by a desire to more strongly differentiate themselves from those who they perceive to be significantly lower status than themselves, those people who are at or close to the bottom rung. This is not surprising given the highly visible and publicly stigmatized status of the poor and homeless in the United States. (Williams, 2009). In this respect, our findings are similar to those found among adults, in which poor participants held more negative views toward welfare recipients than did middle-class participants, and were more likely to rate recipients of welfare as dishonest and idle (Bullock, 1999). The findings are also consistent with a recent study by Flanagan et al. (2014) demonstrating that adolescents from lower SES backgrounds were more likely than those from higher SES backgrounds to make individualistic attributions for poverty. These seemingly contradictory beliefs are consistent with system justification theory, in which people hold beliefs (even when not in their self-interest) that justify the status quo (Jost et al., 2003). Poor adults, for example, are particularly likely to believe that economic inequality is both legitimate and necessary. Perhaps because the status quo provides a sense of security and stability, these system-justifying beliefs, even when the individual is disadvantaged by that system, are shown to be associated with greater satisfaction (Jost et al., 2003).

Study Strengths, Caveats, and Future Directions

This study is important for several reasons. A key strength is the study's mixed-methods approach to investigating children's conceptions about their own and others' social class position. Converging evidence across qualitative and quantitative data sources allowed us to gain a better perspective and insight into children's reasoning about SSS and social class, including how social class informs children's conversations with and social comparisons of themselves vis-à-vis their peers (Creswell & Plano Clark, 2007; Johnson & Onwuegbuzie, 2004). Given that this is the first study to apply a modified version of the SSS ladder measure to preadolescent children, it was important to assess children's comprehension and completion of the task. Furthermore, it was valuable to assess children's perceptions of social class in multiple ways: an open-ended assessment that could, based on what was salient to children, focus on the group, the individual, or the social condition, and a quantitative measure that focused children on group members' attributes. Second, in assessing children's beliefs and stereotypes about social class groups, we included the middle class, in contrast to prior studies that have predominantly focused only on the poor and rich. This inclusion highlighted that children are capable of reasoning about all three distinct social class groups and can rank themselves subjectively in terms of their family's economic position, with a majority showing a preference to rank themselves as middle class.

Third, this study assessed understanding of social class in a diverse setting: at an elementary school where the student body is purposefully diverse with respect to race, ethnicity, SES, and several other dimensions. Because of this diversity, social class may have been more salient for children at this school than for children at more homogenous schools. Our results indicated that children's conceptualization of their SSS was informed by concrete features in their proximal environments, such as how their houses differed from their school peers. Thus, the current study indicates how children *can* understand social class when given a diverse learning environment. Most children in the United States, however, attend schools with little economic diversity; rich children attend schools with other rich children and poor children with other poor children (Orfield & Lee, 2005), and increasingly live in economically segregated communities (Reardon & Bischoff, 2011). Future research should examine whether or not children's assessments are just as accurate in less diverse contexts.

As with all studies, caveats to the study design should be taken into account when interpreting the findings. An important consideration here is the study's correlational design, which cannot provide causal inferences. There is, in general, a paucity of longitudinal research on children's social identity and intergroup attitudes. Longitudinal research is necessary to better understand when children's ideas about social class first begin to develop and the socialization contexts and processes that inform these beliefs. Second, the social group identification measure came before the more evaluative measure and thus may have increased social class group salience in ways that altered the responses to the evaluative measures, a caveat that should be addressed systematically in future studies. A third important caveat to this study is that there were substantially fewer low-SES children in this sample than middle-class and high-SES children. Thus, the responses of children who could be considered poor were not as well represented as those of more affluent children. Additional research should focus more specifically on the conceptualization of SSS for children from low-SES families.

Based on the findings from the current study, we suggest several lines of future research. First, future studies should examine how children learn about social class. For example, research should explore (a) how families talk about sources of family income and wealth, (b) how children's social interactions in and out of school (including extracurricular activities, playdates, and birthday parties that provide opportunities for children and youth to peek into other social worlds of their peers and friends) inform their SSS and beliefs about others, and (c) how the context of neighborhoods and schools influence children's sense of *relative* deprivation or affluence. Recent research has shown that discussing current events and politics with one's parents or attending schools where classmates do so with their families is related to adolescents' beliefs about inequality (Flanagan et al., 2014). The degree to which these sources influence preadolescent children's beliefs about social class remains an open question. Second, having shown that by the end of elementary school, children are capable of reasoning about their own and others' social class groups, research should examine the potential benefits and social costs of group membership, particularly when group membership is as part of a numerical minority in a setting. For example, low-income youth attending high schools with a greater proportion of middle- and high-income students have been shown to fare less well academically and experience more psychosocial problems than those attending schools with more similar SES peers (Crosnoe, 2009). It would be intriguing to assess these relationships on the basis of students' SSS rankings, above and beyond other markers of SES. Only by incorporating both objective and subjective indicators of social class can such questions be answered.

Conclusions

Economic inequality in the United States is at an all-time high and there is little evidence of it reversing course any time soon (Rodrigue & Reeves, 2014; Stiglitz, 2012). Despite volumes of research demonstrating linkages between SES and children's development, little empirical work has investigated whether children's perceptions of their social status influence their outcomes. To guide future research on this topic, the primary aims of the current study were to inform greater understanding of how children think about their SSS and its relation to their beliefs about others' social class group membership. The findings suggest that by the upper elementary school years, children have a relative sense of their own SSS and are able to articulate distinct beliefs about the poor, middle class, and rich in a manner consistent with adults' class-based beliefs. In addition, findings suggest a stronger negativity bias toward the poor among lower SSS children as compared to their more middle SSS peers. Taken together, the results demonstrate that children are not blind to social class. As a society, we need to be mindful that social class not only influences children in the objective sense of what economic resources can provide but also in a more subjective sense through its influence on their developing sense of self and perceptions of others.

References

Aber, L., Morris, P., & Raver, C. (2012). Children, families and poverty: Definitions, trends, emerging science and implications for policy. *Society for Research in Child Development Social Policy Report, 26,* 3–19. Retrieved from http://files.eric.ed.gov/fulltext/ED536953.pdf

Aboud, F. (1988). *Children and prejudice.* New York, NY: Blackwell.

Adler, N., Epel, E. S., Castellazzo, G., & Ickovics, J. R. (2000). Relationship of subjective and objective social status with psychological and physiological functioning: Preliminary data in healthy White women. *Health Psychology, 19,* 586–592. doi:10.1037/0278-6133.19.6.586

American Psychological Association, Task Force on Socioeconomic Status. (2007). *Report of the APA task force on socioeconomic status.* Washington, DC: American Psychological Association.

Bennett, M., & Sani, F. (Eds.) (2004). *The development of the social self.* Hove, UK: Psychology Press.

Bigler, R. S., Averhart, C. J., & Liben, L. S. (2003). Race and the workforce: Occupational status, aspirations, and stereotyping among African American children. *Developmental Psychology, 39,* 572–580. doi:10.1037/0012-1649.39.3.572

Bigler, R. S., Jones, L. C., & Lobliner, D. B. (1997). Social categorization and the formation of intergroup attitudes in children. *Child Development, 68,* 530–543. doi:10.1111/j.1467-8624.1997.tb01956.x

Bigler, R. S., & Liben, L. S. (2006). A developmental intergroup theory of social stereotypes and prejudice. In R. V. Kail (Ed.), *Advances in child development and behavior* (Vol. 34, pp. 39–89). San Diego, CA: Elsevier.

Brown, C. S., & Bigler, R. S. (2005). Children's perceptions of discrimination: A developmental model. *Child Development, 76,* 533–553. doi:10.1111/j.1467-8624.2005.00862.x

Bullock, H. E. (1995). Class acts: Middle-class responses to the poor. In B. Lott & D. Maluso (Eds.), *The social psychology of interpersonal discrimination* (pp. 118–159). New York, NY: Guilford.

Bullock, H. E. (1999). Attributions for poverty: A comparison of middle-class and welfare recipient attitudes. *Journal of Applied Social Psychology, 29,* 2059–2082. doi:10.1111/j.1559-1816.1999.tb02295.x

Chafel, J. A. (1997). Societal images of poverty: Child and adult beliefs. *Youth and Society, 28,* 432–463. doi:10.1177/0044118X97028004003

Chafel, J. A., & Neitzel, C. (2005). Young children's ideas about the nature, causes, justification, and alleviation of poverty. *Early Childhood Research Quarterly, 20,* 433–450. doi:10.1016/j.ecresq.2005.10.004

Cook, T. D., Church, M. B., Ajanaku, S., Shadish, W. R., Jr., Kim, J. R., & Cohen, R. (1996). The development of occupational aspirations and expectations among inner-city boys. *Child Development, 67,* 3368–3385. doi:10.1111/j.1467-8624.1996.tb01918.x

Cook, T. D., & Curtin, T. R. (1987). The mainstream and the underclass: Why are the differences so salient and the similarities so unobtrusive? In J. C. Masters & W. P. Smith (Eds.), *Social comparison, social justice, and relative deprivation* (pp. 217-264). Hillsdale, NJ: Erlbaum.

Creswell, J. W., & Plano Clark, V. L. (2007). *Designing and conducting mixed methods research.* Thousand Oaks, CA: Sage.

Crosnoe, R. (2009). Low-income students and the socioeconomic composition of public high school. *American Sociological Review, 74,* 709–730. doi:10.1177/000312240907400502

Damon, W. (1988). *The moral child.* New York, NY: Free Press.

Diemer, M. A., & Ali, S. R. (2009). Integrating social class into vocational psychology: Theory and practice implications. *Journal of Career Assessment, 17,* 247–265. doi:10.1177/1069072708330462

Diemer, M. A., Mistry, R. S., Wadsworth, M. E., López, I., & Reimers, F. (2013). Best practices in conceptualizing and measuring social class in psychological research. *Analyses of Social Issues and Public Policy, 13,* 77–113. doi:10.1111/asap.12001

Doyle, A. B., & Aboud, F. E. (1995). A longitudinal study of White children's racial prejudice as a social cognitive development. *Merrill-Palmer Quarterly, 41,* 209–228.

DuBois, D. L., Burk-Braxton, C. A., Swenson, L. P., Tevendale, H. D., & Hardesty, J. L. (2002). Race and gender influences on adjustment in early adolescence: Investigation of an integrative model. *Child Development, 73,* 1573–1592. doi:10.1111/1467-8624.00491

Ellemers, N., & van Laar, C. (2010). Individual mobility: The opportunities and challenges members of devalued groups encounter when trying to avoid group-based discrimination. In J. F. Dovidio, M. Hewstone, P. Glick, & V. Esses (Eds.), *The SAGE handbook of prejudice, stereotyping, and discrimination* (pp. 561–576). London, UK: Sage.

Emler, N., & Dickinson, J. (1985). Children's representations of economic inequalities: The effects of social class. *British Journal of Developmental Psychology, 3,* 191–198. doi:10.1111/j.2044-835X.1985.tb00971.x

Evans, M. D., & Kelley, J. (2004). Subjective social location: Data from 21 nations. *International Journal of Public Opinion Research, 16,* 3–38. doi:10.1093/ijpor/16.1.3

Flanagan, C. A., Kim, T., Pykett, A., Finlay, A., Gallay, E. E., & Pancer, M. (2014). Adolescents' theories about economic inequality: Why are some people poor while others are rich? *Developmental Psychology, 50,* 2512–2525. doi:10.1037/a0037934

Flanagan, C., & Tucker, C. (1999). Adolescents' explanations for political issues: Concordance with their views of self and society. *Developmental Psychology, 35,* 1198–1209. doi:10.1037/0012-1649.35.5.1198

Gans, H. J. (1995). *The war against poverty: The underclass and antipoverty policy*. New York, NY: Basic Books.

Goodman, E., Adler, N. E., Kawachi, I., Frazier, A. L., Huang, B., & Colditz, G. A. (2001). Adolescents' perceptions of social status: Development and evaluation of a new indicator. *Pediatrics, 108,* 31–38. doi:10.1542/peds.108.2.e31

Goodman, E., Amick, B. C., Resendes, M. O., Levine, S., Kagan, J., Rogers, W. H., & Tarlov, A. R. (2000). Adolescents' understanding of social class: A comparison of White upper middle-class and working class youth. *Journal of Adolescent Health, 27,* 80–83. doi:10.1016/S1054-139X(99)00116-0

Jahoda, G. (1959). Development of the perception of social differences in children from 6 to 10. *British Journal of Psychology, 50,* 159–175. doi:10.1111/j.2044-8295.1959.tb00693.x

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher, 33,* 14–26. doi:10.3102/0013189X033007014

Jost, J. T., Pelham, B. W., Sheldon, O., & Ni Sullivan, B. (2003). Social inequality and the reduction of ideological dissonance on behalf of the system: Evidence of enhanced system justification among the disadvantaged. *European Journal of Social Psychology, 33,* 13–36.

Katz, M. B. (1989). *The undeserving poor: From the war on poverty to the war on welfare*. New York, NY: Pantheon.

Kotlowitz, A. (1992). *There are no children here: The story of two boys growing up in the other America*. New York, NY: Anchor.

Kozol, J. (1992). *Savage inequalities: Children in America's schools*. New York, NY: Harper Perennial.

Kraus, M. W., Piff, P. K., Mendoza-Denton, R., Rheinschmidt, M. L., & Keltner, D. (2012). Social class, solipsism, and contextualism: How the rich are different from the poor. *Psychological Review, 119,* 546–572. doi:10.1037/a0028756

Ladd, C. E., & Bowman, K. H. (1998). *Attitudes toward economic inequality*. Washington, DC: American Enterprise Institute.

Leahy, R. L. (1981). The development of the conception of economic inequality. I. Descriptions and comparisons of rich and poor people. *Child Development, 52,* 523–532. Retrieved from http://www.jstor.org/stable/1129170

Leahy, R. L. (1983). Development of the conception of economic inequality: II. Explanations, justifications, and concepts of social mobility and change. *Developmental Psychology, 19,* 111–125. doi:10.1037/0012-1649.19.1.111

McLoyd, V., Aikens, N., & Burton, L. (2006). Childhood poverty, policy, and practice. In W. Damon & R. Lerner (Series Eds.) & K. A. Renninger & I. Sigel (Vol. Eds.), *Handbook of child psychology: Vol. 4. Child psychology in practice* (pp. 700–775). Hoboken, NJ: Wiley.

Mistry, R. S., Benner, A. D., Tan, C. S., & Kim, S. Y. (2009). Family economic stress and academic well-being among Chinese-American youth: The influence of adolescents' perceptions of economic strain. *Journal of Family Psychology, 23,* 279–290. doi:10.1037/a0015403

Mistry, R. S., Brown, C. S., Chow, K. A., & Collins, G. S. (2012). Increasing the complexity of young adolescents' beliefs about poverty and inequality: Results of an 8th grade social studies curriculum intervention. *Journal of Youth and Adolescence, 41,* 704–716. doi:10.1007/s10964-011-9699-6

Orfield, G., & Lee, C. (2005). *Why segregation matters: Poverty and educational inequality*. Cambridge, MA: Harvard University, The Civil Rights Project.

Pfeifer, J. H., Ruble, D. N., Bachman, M. A., Alvarez, J. M., Cameron, J. A., & Fuligni, A. J. (2007). Social identities and intergroup bias in immigrant and nonimmigrant children. *Developmental Psychology, 43,* 496–507. doi:10.1037/0012-1649.43.2.496

Quintana, S. M. (1998). Development of children's understanding of ethnicity and race. *Applied and Preventive Psychology: Current Scientific Perspectives, 7,* 27–45. doi:10.1016/S0962-1849(98)80020-6

Ramsey, P. G. (1991). Young children's awareness and understanding of social class differences. *Journal of Genetic Psychology, 152,* 71–82. Retrieved from http://search.proquest.com/docview/1297152026/citation/141CD6DB0AB22B01C41/1?accountid=14512

Reardon, S. F., & Bischoff, K. (2011). Income inequality and income segregation. *American Journal of Sociology, 116,* 1092–1153. doi:10.1086/657114

Rodrigue, E. P., & Reeves, R. V. (2014, October 2). *Horatio Alger goes to Washington: Representation and social mobility*. Retrieved from http://www.brookings.edu/blogs/social-mobility-memos/posts/2014/10/02-horatio-alger-representation-social-mobility-reeves

Ruble, D. N., Alvarez, J., Bachman, M., Cameron, J., Fuligni, A., Coll, C., & Rhee, E. (2004). The development of a sense of "we": The emergence and implications of children's collective identity. In M. Bennett & F. Sani (Eds.), *The development of the social self* (pp. 29–76). New York, NY: Psychology Press.

Sigelman, C. K. (2012). Rich man, poor man: Developmental differences in attributions and perceptions. *Journal of Experimental Child Psychology, 113,* 415–429. doi:10.1016/j.jecp.2012.06.011

Simmons, R. G., & Rosenberg, M. (1971). Functions of children's perceptions of the stratification system. *American Sociological Review, 36,* 235–249. doi:10.2307/2094041

Singh-Manoux, A., Adler, N. E., & Marmot, M. G. (2003). Subjective social status: Its determinants and its association with measures of ill-health in the Whitehall II study. *Social Science & Medicine, 56,* 1321–1333. doi:10.1016/S0277-9536(02)00131-4

Singh-Manoux, A., Marmot, M. G., & Adler, N. E. (2005). Does subjective social status predict health and change in health status better than objective status? *Psychosomatic Medicine, 67,* 855–861. doi:10.1097/01.psy.0000188434.52941.a0

Skafte, D. (1989). The effect of perceived wealth and poverty on adolescents' character judgments. *Journal of*

Social Psychology, 129, 93–99. doi:10.1080/00224545.1989.9711703

Stiglitz, J. E. (2012). *The price of inequality*. New York, NY: Norton.

Tajfel, H. (1982). Social psychology of intergroup relations. *Annual Review of Psychology, 33,* 1–39. doi:10.1146/annurev.ps.33.020182.000245

Trzcinski, E. (2002). Middle school children's perceptions on welfare and poverty: An exploratory, qualitative study. *Journal of Family and Economic Issues, 23,* 339–359. doi:10.1023/A:1021129915436

Tudor, J. F. (1971). The development of class awareness in children. *Social Forces, 49,* 470–476. doi:10.1093/sf/49.3.470

U.S. Census Bureau. (2012). *Statistical abstract of the United States: 2012* (131st ed.). Washington, DC: Author. Retrieved from http://www.census.gov/compendia/statab/

Weinger, S. (1998). Poor children "know their place": Perceptions of poverty, class, and public messages. *Journal of Sociology and Social Welfare, 25,* 100–118.

Weinger, S. (2000). Economic status: Middle class and poor children's views. *Children & Society, 14,* 135–146. doi:10.1111/j.1099-0860.2000.tb00161.x

Williams, W. R. (2009). Struggling with poverty: Implications for theory and policy of increasing research on social class-based stigma. *Analyses of Social Issues and Public Policy, 9,* 37–56. doi:10.1111/j.1530-2415.2009.01184.x

Williams, J. E., Best, D. L., Boswell, D. A., Mattson, L. A., & Graves, D. J. (1975). Preschool Racial Attitude Measure II. *Educational and Psychological Measurement, 35,* 3–18. doi:10.1177/001316447503500101

Woods, T. A., Kurtz-Costes, B., & Rowley, S. J. (2004). The development of stereotypes about the rich and poor: Age, race, and family income differences in beliefs. *Journal of Youth and Adolescence, 34,* 437–445. doi:10.1007/s10964-005-7261-0

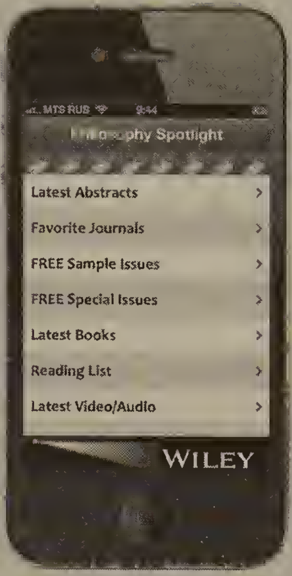
ENABLE
DISCOVERY

WILEY ONLINE LIBRARY
Access this journal and thousands
of other essential resources.

Featuring a clean and easy-to-use interface, this online service delivers intuitive navigation, enhanced discoverability, expanded functionalities, and a range of personalization and alerting options.

Sign up for content alerts and RSS feeds, access full-text, learn more about the journal, find related content, export citations, and click through to references.

WILEY
ONLINE LIBRARY
wileyonlinelibrary.com



MTS RUS
9:44
Philosophy Spotlight
Latest Abstracts
Favorite Journals
FREE Sample Issues
FREE Special Issues
Latest Books
Reading List
Latest Video/Audio
WILEY

WILEY SPOTLIGHT APPS
Download free today

Wiley Spotlight Apps are essential for all researchers, faculty, students and professionals. Whether you want to keep track of broad trends across your discipline or focus on a subfield, the Spotlight Apps are indispensable tools for your research and teaching.

These apps will give you the following features at your fingertips, anytime, anywhere:

- Up-to-the-minute abstracts from leading journals in your field
- Cutting-edge Special and Virtual Issues, with free content
- Latest Video Abstracts, Expert Publishing Workshops, and Informative Podcasts
- Updates and news on key conferences
- And more!

Download Spotlight apps in the following subject areas:

Anthropology
Business
Economics
Education
Geography
History
Linguistics
Philosophy
Politics & Government
Psychology
Religion & Theology
Sociology

www.wileyspotlights.com

WILEY

Save Time and Let the Research Come to You

Sign up for new content alerts for all of your favorite journals:

- ✓ Be the first to read Early View articles
- ✓ Get notified of Accepted Articles when they appear online
- ✓ Receive table of contents details each time a new issue is published
- ✓ Never miss another issue!

Follow these 3 easy steps to register for alerts online:

1 Log into **Wiley Online Library**. If you are not already a registered user, you can create your profile for free.

2 Select "**Get New Content Alerts**" from Journal Tools on the top left menu on any journal page or visit the Publications page to view all titles.

3 Submit your preferences and you are done. You will now receive an email when a new issue of the journal publishes.

12-46060

wileyonlinelibrary.com

REGISTER NOW AT SANTAFEBOYS.ORG

The Psychology of Boys at Risk: Indicators from 0-5

A conference presented by the Santa Fe Boys Educational Foundation in conjunction with the Infant Mental Health Journal

November 5 & 6, 2015
Eldorado Hotel
Santa Fe, New Mexico

SANTA FE BOYS
EDUCATIONAL FOUNDATION

INFANT MENTAL HEALTH JOURNAL
© Michigan Association
for Infant Mental Health

ORGANIZING COMMITTEE

Hiram E. Fitzgerald, PhD
Past Editor-in-Chief
Infant Mental Health Journal

Paul Golding, PhD
Executive Director
Santa Fe Boys Educational Foundation

Paul Spicer, PhD
Editor-in-Chief
Infant Mental Health Journal

Post Office Box 1312
Santa Fe, NM 87504-1312
info@santafeboys.org

Conference Presenters

Panel Discussion follows these presentations. Hiram Fitzgerald, Moderator.
TOPIC: *Bio/Evolutionary Perspectives on Boys at Risk*

KEYNOTE ADDRESS: **Allen Schore, PhD, UCLA**, *The Developmental Neurobiology and Psychology of Boys at Risk*

KEYNOTE ADDRESS: **Melissa Hines, PhD, University of Cambridge**, *Testosterone and the Development of Human Gender-Related Behavior*

Marco Del Giudice, PhD, University of New Mexico, *Attachment and Developmental Risk in Boys: An Evolutionary Perspective*

Panel Discussion follows these presentations. Paul Spicer, Moderator.
TOPIC: *Boys at Risk as Seen Through Longitudinal Outcomes*

Daniel Shaw, PhD, University of Pittsburgh, *Early Childhood Predictors of Boys' Antisocial Behavior and Adjustment in Adolescence and Early Adulthood*

Devi Miron, PhD, Tulane University, *Stress and Psychopathology Following Severe Deprivation: Differential Effects on Boys and Girls*

Marjorie Beeghly, PhD, Wayne State University, *Sex Differences in Early Indicators of Developmental Vulnerability During Early Childhood: Risk, Resilience, and Relationships*

Leon Puttler, PhD, University of Michigan, *Early Risk Factors in the Development of Alcohol and other Substance Use Problems*

Panel Discussion follows these presentations. Marvin McKinney, Moderator.
TOPIC: *Focusing on Cultural Context and Boys at Risk*

Iheoma Iruka, PhD, Buffett Early Childhood Institute, University of Nebraska, *Revisioning Early Childhood Education for Black Boys: Seizing Opportunities and Minimizing Risks*

Natasha Cabrera, PhD, University of Maryland, *The Early Experiences of Latino Boys: A Risk and Protective Approach*

Michelle Sarche, PhD, University of Colorado at Denver, *American Indian and Alaska Native Boys: Risk and Resilience amidst Context and Culture*

Plus Workshops, CEUs, and More!
FULL PROGRAM DETAILS AT SANTAFEBOYS.ORG

Recent

STUDIES IN FETAL BEHAVIOR: REVISITED, RENEWED, AND REIMAGINED
- *Janet A. DiPietra, Kathleen A. Costigan, and Kristin M. Voegtline* (Serial No. 318, 2015).

THE ROLE OF PARENTS IN THE ONTOGENY OF ACHIEVEMENT-RELATED MOTIVATION AND BEHAVIORAL CHOICES
- *Sandra D. Simpkins, Jennifer A. Fredricks, and Jacquelynne S. Eccles* (Serial No. 317, 2015).

SLEEP AND DEVELOPMENT: ADVANCING THEORY AND RESEARCH
- *Mona El-Sheikh and Avi Sadeh* (Serial No. 316, 2015).

THE RELATION OF CHILDHOOD PHYSICAL ACTIVITY TO BRAIN HEALTH, COGNITION, AND SCHOLASTIC ACHIEVEMENT
- *Charles H. Hillman and Carol L. Cheatham* (Serial No. 315, 2014).

THE ADULT ATTACHMENT INTERVIEW: PSYCHOMETRICS, STABILITY AND CHANGE FROM INFANCY, AND DEVELOPMENTAL ORIGINS
- *Cathryn Booth-LaForce, Glenn I. Roisman, John Ruscio, and Marinus H. van IJzendoorn* (Serial No. 314, 2014).

THE EMERGENT EXECUTIVE: A DYNAMIC FIELD THEORY OF THE DEVELOPMENT OF EXECUTIVE FUNCTION
- *Aaron T. Buss, John P. Spencer, Sandra A. Wiebe, and J. Bruce Morton* (Serial No. 313, 2014).

CHILDREN'S UNDERSTANDING OF DEATH: TOWARD A CONTEXTUALIZED AND INTEGRATED ACCOUNT
- *Karl S. Rosengren, Peggy J. Miller, Isabel T. Gutiérrez, Philip I. Chow, Stevie S. Schein, Kathy N. Anderson, and Maureen A. Callanan* (Serial No. 312, 2014).

PHYSICAL ATTRACTIVENESS AND THE ACCUMULATION OF SOCIAL AND HUMAN CAPITAL IN ADOLESCENCE AND YOUNG ADULTHOOD: ASSETS AND DISTRACTIONS
- *Rachel A. Gordon, Robert Crosnoe, and Xue Wang* (Serial No. 311, 2013).

THE FAMILY LIFE PROJECT: AN EPIDEMIOLOGICAL AND DEVELOPMENTAL STUDY OF YOUNG CHILDREN LIVING IN POOR RURAL COMMUNITIES
- *Lynne Vernon-Feagans, Martha Cox, and the FLP Investigators* (Serial No. 310, 2013).

NATIONAL INSTITUTES OF HEALTH TOOLBOX COGNITION BATTERY (NIH TOOLBOX CB): VALIDATION FOR CHILDREN BETWEEN 3 AND 15 YEARS
- *Philip David Zelazo and Patricia J. Bauer* (Serial No. 309, 2013).

RELATIONSHIP PROCESSES AND RESILIENCE IN CHILDREN WITH INCARCERATED PARENTS
- *Julie Poehlmann and J. Mark Eddy* (Serial No. 308, 2013).

THE EMERGENCE OF A TEMPORALLY EXTENDED SELF AND FACTORS THAT CONTRIBUTE TO ITS DEVELOPMENT: FROM THEORETICAL AND EMPIRICAL PERSPECTIVES
- *Mary Lazaridis* (Serial No. 307, 2013).

WHAT MAKES A DIFFERENCE: EARLY HEAD START EVALUATION FINDINGS IN A DEVELOPMENTAL CONTEXT
- *John M. Love, Rachel Chazan-Cohen, Helen Raikes, and Jeanne Brooks-Gunn* (Serial No. 306, 2013).

THE DEVELOPMENT OF MIRROR SELF-RECOGNITION IN DIFFERENT SOCIOCULTURAL CONTEXTS
- *Joscha Kärtner, Heidi Keller, Nandita Chaudhary, and Relindis D. Yovsi* (Serial No. 305, 2012).

"EMOTIONS ARE A WINDOW INTO ONE'S HEART": A QUALITATIVE ANALYSIS OF PARENTAL BELIEFS ABOUT CHILDREN'S EMOTIONS ACROSS THREE ETHNIC GROUPS
- *Alison E. Parker, Amy G. Halberstadt, Julie C. Dunsmore, Greg Townley, Alfred Bryant, Jr., Julie A. Thompson, and Karen S. Beale* (Serial No. 304, 2012).

PHYSIOLOGICAL MEASURES OF EMOTION FROM A DEVELOPMENTAL PERSPECTIVE: STATE OF THE SCIENCE
- *Tracy A. Dennis, Kristin A. Buss, and Paul D. Hastings* (Serial No. 303, 2012).

HOW SOCIALIZATION HAPPENS ON THE GROUND: NARRATIVE PRACTICES AS ALTERNATE SOCIALIZING PATHWAYS IN TAIWANESE AND EUROPEAN-AMERICAN FAMILIES
- *Peggy J. Miller, Heidi Fung, Shumin Lin, Eva Chian-Hui Chen, and Benjamin R. Boldt* (Serial No. 302, 2012).

CHILDREN WITHOUT PERMANENT PARENTS: RESEARCH, PRACTICE, AND POLICY
- *Robert B. McCall, Marinus H. van IJzendoorn, Femmie Juffer, Christina J. Groark, and Victor K. Groza* (Serial No. 301, 2011).

I REMEMBER ME: MNEMONIC SELF-REFERENCE EFFECTS IN PRESCHOOL CHILDREN
- *Josephine Ross, James R. Anderson, and Robin N. Campbell* (Serial No. 300, 2011).

EARLY SOCIAL COGNITION IN THREE CULTURAL CONTEXTS
- *Tara Callaghan, Henrike Moll, Hannes Rakoczy, Felix Warneken, Ulf Liszkowski, Tanya Behne, and Michael Tomasello* (Serial No. 299, 2011).

THE DEVELOPMENT OF AMBIGUOUS FIGURE PERCEPTION
- *Marina C. Wimmer and Martin J. Doherty* (Serial No. 298, 2011).

THE BETTER BEGINNINGS, BETTER FUTURES PROJECT: FINDINGS FROM GRADE 3 TO GRADE 9
- *Ray DeV. Peters et al.* (Serial No. 297, 2010).

FIRST-YEAR MATERNAL EMPLOYMENT AND CHILD DEVELOPMENT IN THE FIRST 7 YEARS
- *Jeanne Brooks-Gunn, Wen-Jui Han, and Jane Waldfogel* (Serial No. 296, 2010).

DEPRIVATION-SPECIFIC PSYCHOLOGICAL PATTERNS: EFFECTS OF INSTITUTIONAL DEPRIVATION
- *Michael Rutter, Edmund J. Sonuga-Barke, Celia Beckett, Jennifer Castle, Jana Kreppner, Robert Kumsta, Wolff Schlotz, Suzanne Stevens, and Christopher A. Bell* (Serial No. 295, 2010).

A DYNAMIC CASCADE MODEL OF THE DEVELOPMENT OF SUBSTANCE-USE ONSET
- *Kenneth A. Dodge, Patrick S. Malone, Jennifer E. Lansford, Shari Miller, Gregory S. Pettit, and John E. Bates* (Serial No. 294, 2009).

FLEXIBILITY IN EARLY VERB USE: EVIDENCE FROM A MULTIPLE-*N* DIARY STUDY
- *Letitia R. Naigles, Erika Hoff, and Donna Vear* (Serial No. 293, 2009).